

C 35, 602 : 80 / 1. 8

# Solar-Terrestrial Predictions Proceedings



Volume 2: Working Group Reports and Reviews

Richard F. Donnelly, Editor

Space Environment Laboratory

Boulder, Colorado



**U.S. DEPARTMENT OF COMMERCE**

National Oceanic and Atmospheric Administration

Environmental Research Laboratories





SOLAR-TERRESTRIAL PREDICTIONS PROCEEDINGS

VOLUME 11

WORKING GROUP REPORTS AND REVIEWS

Edited by

Richard F. Donnelly  
Space Environment Laboratory  
Boulder, Colorado 80303, U.S.A.

December, 1979

The International Solar-Terrestrial Predictions Proceedings and Workshop Program was hosted by the NOAA Space Environment Laboratory. The workshop was held April 23-27, 1979, at the College Inn in Boulder, Colorado.

Science Sponsors of the Program:

AAS: American Astronomical Society  
AGU: American Geophysical Union  
AMS: American Meteorological Society  
COSPAR: Committee on Space Research  
IAGA: International Association of Geomagnetism and Aeronomy  
IAU: International Astronomical Union  
IUWDS: International URSIGRAM and World Days Service  
SCOSTEP: Scientific Committee on Solar-Terrestrial Physics  
URSI: Union Radio Scientifique Internationale; Commissions E and G

Science and Financial Sponsors of the Program:

Air Force Geophysics Laboratory  
Air Force Office of Scientific Research  
Department of Energy  
National Aeronautics and Space Administration  
National Science Foundation  
NOAA Environmental Research Laboratories



## NOTICE

The Environmental Research Laboratories do not approve, recommend, or endorse any proprietary product or proprietary material mentioned in this publication. No reference shall be made to the Environmental Research Laboratories or to this publication furnished by the Environmental Research Laboratories in any advertising or sales promotion which would indicate or imply that the Environmental Research Laboratories approve, recommend, or endorse any proprietary product or proprietary material mentioned herein, or which has as its purpose an intent to cause directly or indirectly the advertised product to be used or purchased because of this Environmental Research Laboratories publication.



## PROGRAM COMMITTEE

F. E. Cook, IPS, Sydney, Australia  
A. D. Danilov, Moscow, U.S.S.R.  
Y. Hakura, Tokyo, Japan  
G. R. Heckman, NOAA SEL, Boulder  
A. P. Mitra, NPL, New Delhi, India  
G. A. Paulikas, Aerospace, Los Angeles  
N. V. Pushkov, IZMIRAN, U.S.S.R.

A. B. Severny, Crimean Observ., U.S.S.R.  
M. A. Shea, AFGL, Bedford, Mass.  
P. Simon, Meudon, France  
W. C. Snoddy, NASA-MSFC, Huntsville  
H. Tanaka, Tokyo Observ., Japan  
J. M. Wilcox, Stanford University  
D. J. Williams, NOAA SEL, Boulder

## WORKING GROUPS and LEADERS

Long-Term Solar Activity Predictions: Mr. P. S. McIntosh, NOAA ERL SEL, Boulder, Colorado 80303 USA

Short-Term Solar Activity Predictions: Dr. P. Simon, Observatory, Meudon, France

Solar Wind and Magnetosphere Interactions: Dr. C. T. Russell, Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

Geomagnetic Storms: Dr. S. I. Akasofu, Geophysical Institute, University of Alaska, College, Alaska 99701, USA

Energetic Particle Disturbances: Dr. G. A. Paulikas, The Aerospace Corporation, P. O. Box 92957, Los Angeles, California 90009, USA

Magnetosphere-Ionosphere Interactions: Dr. Richard R. Vondrak, Radio Physics Lab., SRI International, Menlo Park, California 94025, USA

High-Latitude E- and F-Region Ionospheric Predictions: Dr. Robert D. Hunsucker, Geophysical Institute, University of Alaska, College, Alaska 99701, USA

Midlatitude and Equatorial E- and F-Region Ionospheric Predictions: Dr. Charles M. Rush, Institute of Telecommunications Sciences, National Telecommunications and Information Administration, Boulder, Colorado 80303, USA

D-Region Ionospheric Predictions: Dr. Eivind Thrane, Division for Electronics, Norwegian Defense Research Establishment, P. O. Boks 25, Kjeller, Norway

Solar-Weather Predictions: Dr. K. H. Schatten, Department of Astronomy, Boston University, 725 Commonwealth Avenue, Boston, Massachusetts 02215, USA

Communications Predictions: Dr. A. P. Mitra, Radio Science Division, National Physical Laboratory, Hillside Road, New Delhi - 110012, India

Subsection on Ionosphere-Reflected Propagation: Dr. B. M. Reddy, Radio Science Division, National Physical Laboratory, Hillside Road, New Delhi 110012, India

Subsection on Trans-Ionosphere Propagation: Dr. John A. Klobuchar, AFGL-PHP, Hanscom Air Force Base, Bedford, Massachusetts 01731, USA

Geomagnetic Applications: Dr. Wallace H. Campbell, USGS Box 25046, Denver Federal Center MS 946, Denver, Colorado 80225, USA

Space-Craft Environment and Manned Space Flight Applications: Dr. A. L. Vampola, The Aerospace Corporation, P.O. Box 92957, Los Angeles, California 90009, USA





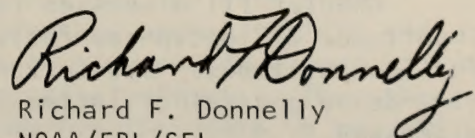
## PREFACE

The International Solar-Terrestrial Prediction Proceedings and Workshop Program (ISTP/P-W Program) included the following: (1) an open call for contributed papers on solar-terrestrial predictions; (2) invited review papers about (a) the prediction, warning and monitoring services of groups that regularly issue solar-terrestrial predictions; (b) the current and future needs for predictions by groups that use solar-terrestrial predictions, and (c) current knowledge of selected topics in solar-terrestrial physics and applications; (3) working groups on fourteen areas of interest for solar-terrestrial predictions; (4) a preprint exchange from October, 1978 through March, 1979; (5) a workshop of representatives of the working groups; and (6) the Solar-Terrestrial Predictions Proceedings. These proceedings consist of four volumes:

Volume I.	Prediction Group Reports (GPO No. 003-023-00041-9)
Volume II.	Working Group Reports and Reviews
Volume III.	Solar Activity Predictions
Volume IV.	Prediction of Terrestrial Effects of Solar Activity

Volume I reviews the current practice in solar-terrestrial predictions. Volume II presents the recommendations and reports developed by the working groups at the workshop. Topical reviews and papers on the current and future needs for predictions are also included. Volumes III and IV present individual suggestions for particular prediction schemes.

The goals of the program were as follows: (1) to determine and document the current state-of-the-art of solar-terrestrial predictions, the applications of these predictions, and the future needs for solar-terrestrial predictions, (2) to encourage research, development and evaluation of solar-terrestrial predictions, and (3) to provide indepth interaction of prediction users, forecasters and scientists involved in the research and development of prediction techniques. To achieve the first goal, we invited forecast groups and user groups to review their activities. The working groups concentrated on deriving recommendations for future needs pertinent to solar-terrestrial predictions. The early call for contributed papers was made to achieve the second goal, i.e. authors had more than a year to orient their work towards a paper on predictions. The workshop was aimed at the third goal. The criteria used for selecting working-group representatives to attend the workshop included the goal of having at least one forecaster per group. The working groups on solar-activity, magnetospheric physics, ionospheric physics, etc., were aimed at attracting scientists. On the other hand, the working groups on communications, geomagnetic applications, and space-craft and man-in-space applications were intended to attract prediction users as well as applications scientists.

  
Richard F. Donnelly

NOAA/ERL/SEL  
Boulder, Colorado 80303 USA

December 3, 1979



## OVERVIEW

Volume II of the Solar-Terrestrial Prediction Proceedings presents the main results of the workshop and the recommendations of the working groups. Three types of papers are included: (1) working group reports, (2) topical review papers and (3) reviews of the needs for predictions by groups that use solar-terrestrial predictions. In some cases, papers originally submitted by their authors as contributed papers were included in this volume as topical reviews. Note that the table of contents lists the location of specific working group recommendations.

The world-wide group of solar-terrestrial prediction services discussed in Volume I of these proceedings is currently dominated by radio communication predictions. Volume II strongly emphasizes predictions for satellite and power-system applications because of their increasing importance. Similarly, transionosphere radio-propagation is stressed because of its increasing importance relative to ionosphere-reflected communications. Furthermore, the opportunities for improving prediction services in the next decade are probably greater for these newer areas of applications than in the traditional ionosphere-reflected radio communications predictions. On the other hand, many of the problems in satellites, long-line power systems or pipelines caused by solar activity will be minimized through improved engineering designs that reduce the sensitivity of future systems to solar-terrestrial disturbances; the application of solar-terrestrial predictions to help solve these problems is only one of several approaches. For example, improved engineering design has satisfactorily solved the solar-terrestrial disturbance problems in long-line cable communications.

Chapter I of this volume discusses in depth the solar-terrestrial environment problems for spacecraft, including what predictions are required for energetic particles, the low energy plasma, and the neutral atmosphere. Our current knowledge of these three topics and the effects of solar-terrestrial disturbances on manned space flight are reviewed.

Recent trends toward interconnecting electric power systems over large geographic areas, increasing development of large electric power systems at high latitudes and opening of oil or gas fields with long pipelines at high latitudes have led to increased importance of predicting geomagnetic activity and its ground-induced currents. Chapter II discusses the needs for solar-terrestrial predictions and further research for this exciting and rapidly developing applications area that places new emphasis on predicting the time rate of change of the geomagnetic field in particular geographic regions.

Chapter III discusses radio communication predictions. The continued importance of ionosphere-reflected communications is reaffirmed, especially for low-latitudes. The current state-of-the-art of modeling the ionospheric time-delay and scintillation effects on transionosphere radio propagation is discussed.

Chapter IV discusses solar-activity predictions. Recently, interest in long-term solar-activity variations has been renewed with wider acceptance



of its possible effects on climate and the occurrence of phenomena like the Maunder Minimum. In the coming years, worldwide shortages in energy, food, and raw materials will increase the importance of long-term predictions. The variety of published predictions for the current sunspot cycle includes such a large range of peak values that it is clear that the theories upon which many of these prediction schemes are based must be inconsistent with the physical processes governing the sunspot cycle. Using sunspot data during a cycle to predict the sunspot number about one year in advance produces sufficiently accurate results. Using data at sunspot minimum to predict the following sunspot cycle provides useful results. Recently Ohl in the U.S.S.R., Brown in England, and Sargent in the United States have each used the effects of recurrent geomagnetic activity during the decline of one eleven-year cycle to predict the following sunspot cycle. They provide the first major advance in long term solar activity predictions beyond the state-of-the-art given in Yu. I. Vitinskii's book Solar-Activity Forecasting (1962, Academy of Sciences of the U.S.S.R., Leningrad; English translation NASA TT F-289, TT65-50115, 1965). On the otherhand, these recent improvements still do not provide predictions far enough in advance to satisfy many system designers who would like to have good predictions 15 to 30 years in advance. This is the time frame where the gap between the prediction needs and current prediction performance is largest with little hope of improvement in the near future. The statistical techniques used for 15 to 30 year predictions are based on data that do not include the Maunder Minimum. If several proposed very long-term periodicities in sunspot numbers are correct, they should contribute to low sunspot numbers during cycle 22. We need to better understand the physical mechanisms that cause sunspot cycles in order to improve our predictions.

The main improvements in short-term predictions in recent years have been the improved real-time monitoring of the sun at radio, X-ray and optical wavelengths, the detection and diagnosis of solar flares, and the short-term prediction first of energetic particles and then of magnetospheric and ionospheric storms. Predictions of major solar flares several days in advance depend strongly on active region evolution studies using optical measurements with high spatial resolution. Recent research on coronal holes indicates that the addition of solar X-ray or EUV monitoring measurements with moderate spatial resolution using existing technology should provide improved predictions of sector boundary passages and geomagnetic activity.

Chapter V discusses predictions of magnetospheric disturbances including solar-wind and magnetosphere interactions. One of the main difficulties in predicting the terrestrial effects of solar activity has been that the traditional geomagnetic activity indices do not provide accurate quantitative indication of the strength of particular terrestrial effects of geomagnetic activity, e.g. the strength of induced ground currents or the decrease in an ionospheric parameter like foF2. A major improvement may now have broken this impasse; namely, measurements of the interplanetary magnetic field and solar wind parameter upstream from the magnetosphere may provide a good indicator and predictor of geomagnetic and other terrestrial effects. The paper by Tsurutani and Baker suggests a practical program for ISEE data to be used to achieve this in real time. In effect, the problem of improving predictions of geomagnetic activity and its associated terrestrial effects may be solved by bypassing the traditional geomagnetic indices.

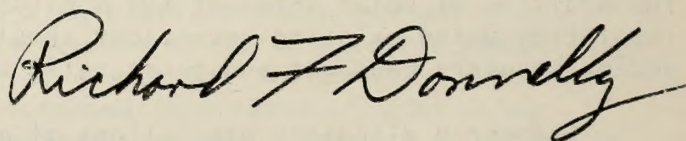


Chapter VI discusses ionospheric predictions. It emphasizes the scientist's viewpoint that improved understanding of the physical processes that link solar and magnetospheric disturbances with the various ionospheric effects. This viewpoint contrasts somewhat with many forecastors of ionosphere-reflected radio communications. They have waited many years for such improvements and in the meantime have had far more success with empirical and statistical approaches. Perhaps the expected improvements in predicting magnetospheric and geomagnetic disturbances will also finally provide the long awaited improvements in ionospheric predictions and radio communications predictions.

The final chapter discusses the potential for predicting climate or weather effects of solar activity. We do not yet know the physical mechanisms that link solar activity with climate. We only have studies of statistical correlations that suggest that solar activity may cause subtle changes in the climate. In effect, the subject of sun-weather relations has only recently risen slightly above the threshold level of "statistical significance". Numerous qualitative theories have been proposed for the physical mechanisms involved.

It may even be possible that both solar activity and climate are responses to some other third source of variations, e.g. galactic gravity waves. The tedious task of examining and testing the possible mechanisms has begun. This field must now concentrate on basic research. It is too premature to issue climate predictions based on variations of solar activity. On the otherhand, predictions in other areas where the physical mechanisms are known still tend to be based on empirical relations derived from correlative studies. However, the correlation coefficients in sun-weather relations are usually not very high and therefore do not lend themselves to accurate predictions. Chapter VII explores how the field of sun-weather research should consider predictions.

Overall, the workshop was quite successful, due largely to the efforts of the working group leaders. I recommend that any subsequent workshop on solar-terrestrial predictions concentrate on particular prediction problems and include data and comparisons of forecasts for preselected periods of solar activity.



Richard F. Donnelly  
ISTP/P-W Program Chairman

#### ACKNOWLEDGEMENTS

The editor wishes to thank the many persons and organizations who helped conduct or support the International Solar-Terrestrial Predictions Proceedings and Workshop Program. Without their support, this program would not have been possible. I am grateful for the science sponsors and financial sponsors listed on the title page. I appreciate the extensive work of the Working Group Leaders listed on the page after the title page. Their work made the workshop a success and created most of the contents of this volume.



# SOLAR-TERRESTRIAL PREDICTIONS PROCEEDINGS

## VOLUME II WORKING GROUP REPORTS AND REVIEWS

### TABLE OF CONTENTS

Program Committee and Working Group Leaders.....	iii
Preface.....	v
Overview.....	vi
Acknowledgements.....	viii
 I. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS — A.L. Vampola et al.....	 1
1. Introduction.....	1
2. User Requirements of Solar-Terrestrial Predictions for Spacecraft Applications: <u>Energetic Particles</u> .....	4
3. User Requirements of Solar-Terrestrial Predictions for Spacecraft Applications: <u>Plasma</u> .....	9
4. User Requirements of Solar-Terrestrial Predictions for Spacecraft Applications: <u>Neutral Atmosphere</u> .....	11
5. Recommendations.....	16
6. Minority Report.....	18
7. References.....	19
 Modeling the Earth's Radiation Belts — J.I. Vette, M.J. Teague, D.M. Sawyer and K.W. Chan.....	 21
 Low Energy Magnetospheric Plasma Interactions with Space Systems - The Role of Predictions — H.B. Garrett.....	 36
 Predictability of Upper-Atmospheric Density and Composition — J.M. Straus and D.R. Hickman.....	 71
 Solar-Induced Variations of Environment Affecting Manned Space Flight and Spacecraft Operations — W.G. Johnson, T.A. Parnell, W.W. Vaughan.....	 89
 Prediction of Spacecraft Potentials at Geosynchronous Orbit — H.B. Garrett, A.G. Rubin and C.P. Pike.....	 104
 The Terrestrial Radiation Environment and EVA'S: Prediction Requirements, Model Improvements, and Warning Systems — E.G. Stassinopoulos.....	 119
 Dosimetry in the Manned Space Program — W. Atwell.....	 121
 II. GEOMAGNETIC APPLICATIONS — W.H. Campbell et al.....	 133
1. Electric Power Transmission.....	133
2. Oil and Gas Pipelines.....	134
3. Magnetic Mapping for Geological and Navigational Requirements....	135

4. Prediction Needs for Geomagnetic Applications.....	136
5. Association on the Effects of Geomagnetically Induced Currents...	136
Magnetic Storm Effects in Electric Power Systems and Predictions Needs — V.D. Albertson and J.G. Kappenman.....	137
The Problem of Solar Induced Currents — D.H. Boteler.....	149
Prediction of Solar Induced Currents and Effects on Power Transmission Systems in Central Canada — W.R. Goddard and W.M. Boerner.....	162
Harmonics of 60 Hz in Power Systems Caused by Geomagnetic Disturbances — K. Hayashi, T. Oguti, T. Watanabe, K. Tsuruda, S. Kokubun and R.E. Horita.....	172
Induced Electric Currents in the Alaska Oil Pipeline Measured by Gradient, Fluxgate, and Squid Magnetometers — W.H. Campbell and J.E. Zimmerman.....	182
Very Large Geomagnetic Disturbances During Sunspot Cycle 21 - A Prediction — H.H. Sargent III.....	193
Potential Prediction Needs in Support of Energy Systems — R.L. Blake.....	198
III. COMMUNICATIONS PREDICTIONS — A.P. Mitra.....	203
A. Ionosphere-Reflected Propagation — B.M. Reddy.....	203
Recommendations.....	211
B. Transionospheric Propagation Predictions — J. Klobuchar.....	217
1. Introduction.....	217
2. Time Delay Predictions.....	220
3. Scintillation Modeling.....	228
4. Conclusions and Recommendations.....	237
IV. SOLAR-ACTIVITY PREDICTIONS	
A. Long-Term Solar Activity Predictions — P.S. McIntosh.....	246
Recommendations.....	254
A New Method of Very Long-Term Prediction of Solar Activity — A.I. Ohl and G.I. Ohl.....	258
New Methods for Predicting the Magnitude of Sunspot Maximum — G.M. Brown.....	264
The Variability of the Solar Ultraviolet Spectral Irradiances — J.G. Timothy.....	280



B.	Short-Term Solar Activity Predictions — P. Simon.....	287
	Conclusions and Recommendations.....	315
	The Needs for Predictions and Real-Time Monitoring for the Flare Build-Up Study — Z. Svestka.....	322
	Solar Forecast and Real-Time Monitoring Needs of the Study of Energy Release in Flares (SERF) — D.M. Rust.....	331
	Requirements for Predictions and Real-Time Monitoring for the Study of Travelling Interplanetary Phenomena — M. Dryer and M.A. Shea.....	340
V.	MAGNETOSPHERIC PREDICTIONS	
A.	Solar Wind and Magnetosphere Interactions — C. Russell et al....	346
	Recommendations.....	357
	On the Prediction of Magnetospheric Configuration — J.A. Slavin and R.E. Holzer.....	365
	The Magnetospheric Electric Field and Convective Processes as Diagnostics of the IMF and Solar Wind — S.M. Kaye.....	375
	Review of Selected Geomagnetic Activity Indices — J.H. Allen and J. Feynman.....	385
	The Prediction of AE, ap, and Dst at Time Lags Between 0 and 30 Hours — D.F. Smart et al.....	399
B.	Prediction of Energetic Particle Disturbances — G. Paulikas et al.....	415
	Recommendations.....	416, 426
	Appendix I. Short Term Magnetospheric Particle Variations (1 min < T < 1 day) — P.R. Higbie et al.....	433
	Appendix II. Long Term Magnetospheric Particle Variations (1 Day < T < ∞ ) — J.I. Vette et al.....	441
	Energetic Solar Particle Behaviour in the Magnetosphere — M. Scholer.....	446
	A Real-Time ISEE Data System — B.T. Tsurutani and D.N. Baker....	464
VI.	IONOSPHERIC PREDICTIONS	
A.	Magnetosphere-Ionosphere Interactions — R.R. Vondrak et al.....	476
	Recommendations.....	491

Auroral Magnetosphere-Ionosphere Coupling: A Brief Topical Review — Y.T. Chiu, J.M. Cornwall and M. Schulz.....	494
B. High-Latitude E- and F-Region Ionospheric Predictions — R.D. Hunsucker et al.....	513
Recommendations.....	522
On the Approach to the Forecast of the Polar Ionosphere Conditions — A.S. Besprozvannaya, A.V. Shirochkov, and T.I. Shchuka.....	528
Morphology and Phenomenology of the High-Latitude E- and F- Regions — R.D. Hunsucker.....	543
C. Report of the Mid- and Low-Latitude E and F Region Working Group — C. Rush.....	562
Recommendations.....	568
D. D-Region Predictions — E.V. Thrane et al.....	573
Recommendations.....	573
Ion Production in the D-Region — W. Swider.....	599
Effects of D-Region Ionization on Radio Wave Propagation — T.R. Larsen.....	617
VII. SOLAR WEATHER/CLIMATE PREDICTIONS — K. Schatten et al.....	655
1. Overview.....	655
2. Objectivity and Statistical Problems.....	656
3. Intermediary Mechanisms Between the Sun and Weather.....	658
4. Solar and Terrestrial Parameters.....	659
5. Specific Predictions and Further Studies.....	662
6. Recommendations.....	666
Coupling Processes Related to the Sun-Weather Problem -- R.A. Goldberg and J.R. Herman.....	669
The Impact of Sun-Weather Research on Forecasting — M.F. Larsen.....	689
Planetary Waves and Sun-Weather Effects — H. Volland and J. Schafer.....	703
The Clouded Crystal Ball: Comments on Geophysical Prediction — S.M. Silverman.....	722



# I. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS

A WORKING GROUP REPORT prepared by: A. L. Vampola, Chairman, R. Altrock, W. Atwell, R. C. Backstrom, H. B. Garrett, D. Hickman, R. G. Johnson, N. H. Pereyaslova, D. Robbins, J. Slowey, J. Sroga, E. Stassinopoulos, M. Teague, and W. W. Vaughan

## 1. INTRODUCTION

Our working group limited its scope to those areas which are influenced by solar-terrestrial coupling effects and are internal to the earth's magnetosphere. We did not consider, for instance, high altitude aircraft in the polar regions (exposed to solar protons at times of large flares) or missions to the moon or other planets, although they, too, are exposed to significant fluxes of energetic solar protons following proton flares. However, factors considered within the limited scope do have other applications and may address the above areas, although indirectly. Three subgroups were established which considered the areas of Energetic Particles, Plasma, and Neutral Atmosphere. Subsequent sections of this report deal with the requirements of users, the current status of models and predictive techniques, and future areas for research in each of these categories. Before we go into the detailed discussions however, we shall briefly discuss the types of users to be considered and the types of phenomena involved.

### 1.1 Users

Users of predictive techniques can be classified by the immediacy of their requirements. Operational programs, i.e., those which are currently engaged in a mission with spacecraft (manned or unmanned) already in orbit, have the highest requirement for accurate short and medium range predictions (24 hrs - 1 yr). Manned operations require real-time monitoring of solar parameters to deal with the possible threat of solar proton flares and magnetic field parameters to predict large scale energetic electron acceleration by major magnetic storms. In the event of a flare, predictive techniques must be available which, with high confidence, can detail the evolution of particle distributions at the spacecraft for minutes to hours and perhaps days ahead. If large space structures at synchronous altitude are serviced by men, the predictive techniques would have to extend to the evolution of the energetic particle population in the outer magnetosphere during and following magnetic storms. Other manned missions have similar requirements for other parameters. Another class of missions which require real-time monitoring and short time predictive techniques is that which includes very low altitude satellite orbits in which the atmospheric heating response to solar events produces significant increases in drag. For virtually every parameter under consideration, a mission can be envisioned which would require real-time or near-real-time monitoring of the parameter and accurate predictive techniques describing the temporal and/or spatial evolution of that parameter.

Presently, the operations people have limited predictive capability. Indeed they are forced to operate most often in the reaction mode and answer such questions as: a) How long will a proton or plasma event last? b) What is the intensity of the event? c) What effects will the event have on a particular system?

The second class of users is that which encompasses those who require models or probabilistic predictions of a longer duration. Spacecraft hardware designers are an example. After a mission has been defined, i.e., orbit, launch date and flight duration are known, engineers require the tools to design a system which will survive and operate in the environment. An approach which uses either a worst case specification or an average-expected-environment-with-safety-factor model is usually satisfactory. In this instance, the requirement is not really for a predictive technique but rather an understanding of the upper and lower limits which a parameter may be expected to assume. Magnetospheric radiation belt modeling is an example of this approach.

A third category of user is the mission planner. He requires both a modeling capability and medium to long range predictions of solar effects. His selection of orbit and mission duration are governed to some extent by the typical environment which may be encountered, provided he knows sufficiently well what that environment might be. His selection of launch date (and perhaps mission duration) will depend upon his ability to predict the specific environment to be encountered at the time of launch. Manned missions, for instance, might not be launched at a time when a co-rotating solar proton source region was active on the sun if the mission used either a terrestrial polar orbit or one going beyond the magnetosphere.

## 1.2 Phenomena

Here we can discuss only those phenomena which we now believe to have deleterious effects upon spacecraft, their missions, and their occupants. From time to time new areas of concern are identified, either because the type of mission or orbit is new or a new technology is used which is susceptible to disturbance by a parameter which was previously thought to be either inconsequential or benign. The latest example is the 'soft failure' of logic elements in microminiaturized electronics. This will be discussed below.

### 1.2.1 Dose.

By 'dose' we refer to all phenomena in which it is the cumulative effect of ionization within an element (human cell, semiconductor junction) which causes damage. A related effect is the accumulation of free charge within insulators (i.e., cable dielectrics) and subsequent arcing which produces transients within the spacecraft electronics. The period of accumulation may be very long, as in communications satellites which have a planned life expectancy of seven to ten years, or very short, as in the case of an astronaut on an EVA (Extra-Vehicular Activity) at a time when the natural environment might subject him to a skin dose rate of several hundred to a thousand rems per hour. The specific case of dose considerations in manned missions is addressed in a short contribution by Atwell (1979), Stassinopoulos (1979) and briefly by Johnson et al. (1979) in these proceedings. We should point out that greater accuracy is required for calculations of dose for manned missions than for unmanned, since there is less variation in dose tolerance and the consequences are more critical in man than in components.

### 1.2.2 Charging.

In the presence of a 'hot' plasma, a spacecraft can charge to significant negative potentials due to the much greater mobility of the electrons than the ionic constituent of the plasma. The resulting charge on the spacecraft can lead to a number of



deleterious effects, some of them potentially lethal to the hardware. The negative potential may enhance the accumulation of contaminants by spacecraft surfaces through the mechanism of outgassing, photoionization of the outgassed molecule, and re-attraction by the potential on the spacecraft. Thermal control surfaces, especially second-surface mirrors, optical devices and certain sensors are particularly vulnerable to this type of degradation. If the spacecraft is in sunlight, photoemission of electrons from the sunlit portions of the vehicle may neutralize the charge build-up due to the plasma. The self-shadowed portions of the vehicle, unless they are electrically conducting and connected to a sunlit conductor, will remain at high negative potential. The potential difference between these differentially charged surfaces may result in an arc which could interfere with the operation of the vehicle or even destroy a critical component. An EVA at synchronous orbit could result in some rather spectacular (and perhaps tragic) events if this phenomenon is not considered in the design of the spacesuit and the EVA itself. Perhaps real-time monitoring of electron pitch-angle distributions (see Baker et al., 1979) will be required to predict the onset of substorms (and their injection/acceleration of hot plasma into the region of the large space structure being serviced) if the synchronous orbit becomes populated by structures serviced by man. The charging phenomenon is not now well understood, but the SCATHA (Spacecraft Charging at High Altitude) program, with the quasi-synchronous spacecraft P78-2, is now investigating the problem both in the lab and in space and will perhaps eventually provide both an understanding of the phenomenon and quantitative models for predictive purposes.

### 1.2.3 Neutral Atmosphere

An important use of models of the neutral atmosphere above 100 km has been in the general area of prediction of satellite orbit evolution. The upcoming demise of Skylab has focussed public attention on both the importance and limitation of our ability to predict satellite drag. The economic cost of inaccurate orbit decay predictions can be substantial. Other important uses of the models can be found in such areas as the planning of scientific missions; the interpretation, reduction, and analysis of experimental results; and the development of models relating the neutral upper atmosphere to the ionosphere and magnetosphere.

The MSIS model (Hedin et al., 1977a, 1977b) and the most recent Jacchia model (Jacchia, 1977) are both capable of representing the density and composition of the thermosphere and exosphere with an accuracy approaching that of the available observational data, at least for time-scales of a day or more. Of these two models, the Jacchia model represents the geomagnetic variation more accurately, while the MSIS model represents the diurnal variations more accurately in the lower thermosphere. Both models can be improved in these two areas and work is in progress to accomplish this. The CIRA 1972 model (COSPAR, 1972) continues to be useful in providing estimates of total atmospheric density for heights above 200 km. However, this and other older models do not correctly represent the strong dependence of the geomagnetic variations on geomagnetic latitude.

Current neutral atmosphere models reflect the available measurements very well and use the  $F_{10.7}$  and  $A_p$  or  $K_p$  indices to indicate the amount of energy input into the thermosphere. If present neutral density models are to be used in making predictions, therefore, there must exist a significant ability to predict these indices. There is only very limited current capability in predicting these indices and we have not seen much promise for greatly improved prediction ability in the near future.



It is clear that even with improved capability to predict  $A_p$  and  $F_{10.7}$ , substantial model improvement must depend on utilization of parameters which more directly measure the thermospheric energy sources. If models which use these improved parameters are to be developed it is essential to ensure that future satellite missions simultaneously measure the energy inputs and the ensuing atmospheric response.

#### 1.2.4 Soft Failures

The increasing use of very high density micrologic in spacecraft systems has introduced a new concern of solar influences. A relatively low energy ( $\approx 2$  MeV/nucleon) high-Z (Fe, for instance) cosmic ray has sufficient specific ionization rate that the energy deposited in a microjunction can cause a change of state of the device (a 'bit-flip'). In some cases, the result may be a 'latch-up' which destroys the device, but usually the only result is a bit error in the spacecraft logic. The error may be unnoticeable in the data or may result in loss of a spacecraft function. Typical results for some tested devices range from no susceptibility to one bit flip per bit per  $10^4$  days (equivalent to 4 bit-errors per hour for a megabit memory). This phenomenon is discussed by Sivo et al. (1979) and test results are presented by Kolasinski et al. (1979). A prediction capability for solar flares rich in high-Z nuclei should be developed to cope with this effect.

#### 1.2.5 Background/Interference

These effects are due to the same type of interaction discussed under Dose and Charging, but are distinguished from them in that they result in no permanent damage to a system. They merely degrade data or make a spacecraft subsystem temporarily inoperable. These subsystems usually include a sensitive detector of one sort or another. The famous star-sensor that lost lock on Canopus every time an energetic proton mimicked the light output from the star is a good example. Another is the X-ray telescope that was swamped by background counts each time it went through the South Atlantic Anomaly. Usually, sensors are designed to work in spite of these background effects; however, the environment and variations in it must be accurately predicted in order to do the necessary design.

## 2. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS: ENERGETIC PARTICLES

Prepared by the Energetic Particles Subgroup: A. L. Vampola, Chairman, W. Atwell, R. C. Backstrom, R. G. Johnson, N. H. Pereyaslova, D. Robbins, J. Sroga, E. Stassinopoulos, and M. Teage

In this section we consider all particles which produce deleterious effects through energy deposition. In general, electrons below 50 keV and protons below 100 keV are not in this category. However, for certain applications, these energies are sufficient to cause direct damage; in those instances, we include them in our discussion of predictive techniques. Modeling is considered one form of prediction and will be treated as such.

We shall subdivide the energetic particle discussion into categories based on two distinct parameters: Physical location and particle energy/type. The requirements for predictive techniques and the state of the art differ markedly for the inner magnetosphere, the outer magnetosphere, low polar orbit, interplanetary trajectories, and

planetary encounters. Low polar orbit and interplanetary trajectories share a common concern for solar proton flux predictions. We will not address planetary encounters. Basically, particles will be divided up into electrons and protons (and other ions); energies will be low (up to 500 keV for electrons and 5 MeV for protons), medium (to 2 MeV for electrons and 50 MeV for protons) and high (everything above medium). We will also consider particle origin and transport as needed.

## 2.1 Magnetospheric Zone

### 2.1.1 Inner Zone

For our purposes we shall consider the inner zone to be all altitudes and latitudes below  $L=2$ . ( $L$  is McIlwain's parameter and in a dipole field corresponds to the radial distance from the center of the earth to the equatorial crossing of a given field line. Units are in earth radii.) The effects which are important in this region are dose effects to man and components, background effects in sensors, and transient upsets in logic due to high- $Z$  events as discussed in the introduction. If we initially limit our discussion to low inclination orbits, we don't have to consider solar and galactic cosmic rays. Presently, there is no knowledge of significant fluxes of high- $Z$  nuclei with energies in the MeV per nucleon range in the inner zone. For polar orbits, we will have to consider these particles.

The prediction of particle fluxes in the inner zone is in excellent shape with the radiation models issued by the National Space Science Data Center (see the review by Vette et al., 1979). The modeling of energetic protons in the inner zone up to hundreds of MeV has progressed to the point that the prediction is probably as reliable as any given single measurement of the flux. At the lower altitudes ( $< 1000$  km) the latest models include solar-cycle effects. The prediction of dose due to energetic protons under thick shields is certainly better than a factor of two and is probably within the 25% range for reasonably long duration missions. The low energy proton population ( $< 5$  MeV) is subject to significant temporal variations but the cumulative effects from this portion of the particle environment are small compared to those from the higher energy population. To summarize the state of predictive capability in the inner zone proton population, the models are adequate for all present missions and may be presumed to be correct unless new reliable data are obtained to the contrary. The weakest area of knowledge is the regime covering the energy range above a few hundred MeV. Spectra, pitch-angle distributions, and flux intensities could be used in sensor design and background estimation if they became available. However, we know of no present or future mission for which such information would provide primary design criteria.

In the inner zone, the electron modeling situation is also excellent. Fluxes above 1 MeV are sufficiently low at all times that for most purposes they can be ignored. Substantial fluxes of lower energy electrons are subject to significant variations only on time periods commensurate with the solar cycle. At times of very large magnetic storms, a small fraction of the energetic outer zone flux succeeds in diffusing through the slot region. Electrons with energies up to 1.5 MeV have been traced in to as far as  $L=1.55$ . But the contribution of these particles to the total dose in inner-zone orbits is negligible. They can constitute a significant increase in background for some sensor systems. Prediction of such events would be useful but presently is not essential. Any predictive technique would probably require as inputs a knowledge of the energetic electron population in the outer zone and a detailed knowledge of the behavior of the



earth's magnetic field during the diffusion period (since it is the magnetic behavior that is driving the diffusion). The outer zone electron population is itself a response to earlier magnetic field behavior. Predictive techniques which are required to provide a knowledge of the outer zone electron flux will be discussed later. At our present state of knowledge of interactions between the solar wind and the magnetosphere, it is unlikely that we will be able to predict, in the foreseeable future, geomagnetic activity with sufficient accuracy to enable a prediction of energetic electron transport into the inner zone. (The above statement is partially based on our current lack of a detailed understanding of loss processes in the slot region.) It should be pointed out that natural sources of electrons with energies above 1 MeV would supply orders of magnitude less flux in the inner zone than was injected by the Starfish nuclear event. Spacecraft and sensors were designed to be operable in the remnants of that injection during the mid-Sixties and could again be made so.

To summarize for the inner-zone orbits: Models are currently adequate and constitute all of the predictive techniques currently required. A better understanding of particle transport through the slot region into the inner zone would, in special cases, be useful, providing that the outer zone electron population were known at the same time. This indicates a need for either a sophisticated prediction scheme for outer zone electron acceleration and transport or a real-time monitoring system. The utility of the data for inner-zone purposes does not warrant the expenditure necessary to obtain it. (There are strong reasons for obtaining the data for outer zone missions; hence, the data may be available essentially 'free'.)

#### 2.1.2 Low Altitude Polar

For polar orbits, all of the above considerations apply, since a satellite in low polar orbit spends about 40% of its time in the inner zone. It also spends about 25% of the time in the outer zone and 35% in polar regions. This could be considered a 'composite' orbit which includes aspects of the inner zone, the outer zone, and interplanetary missions. The additional requirements for this orbit can be obtained from the following sections which address the outer zone and interplanetary missions.

#### 2.1.3 Outer Zone

The situation in the outer zone is much more complex than in the inner zone. The principal reason is that it is much more dynamic. The electron component has a relatively short residence time in the magnetosphere (decay times of the order of a few to tens of days) but the source activity has commensurate time periods. Sources may be particles injected from the tail coupled with radial diffusion as an accelerator or in-situ acceleration by non-adiabatic means. Energetic electrons in the inner region of the outer magnetosphere can change by orders of magnitude in flux level over periods of less than a day. (See Vette et al., 1979). The situation with regard to protons is somewhat better.

From the point of view of the user, the primary areas of interest are the variability of the energetic electron fluxes and the access of solar protons. Solar protons do not constitute a significant portion of the dose to a spacecraft for a long term mission except for interplanetary missions. However, for short periods, they can constitute a sizeable portion of the dose-rate. The predictive capability which is needed in this area is the following: Given a solar flare on the sun, predict the fluence and spectrum as a function of time and L within the magnetosphere. A desired goal is the capability to predict the flare, itself, from the features on the sun and the sun's

immediate past behavior. Both of these predictive capabilities may still lie far in the future. However, an immediate goal should be to obtain sufficient understanding of the physics of the entry process to be able to perform the first stated prediction once a measurement of the proton flux has been made in interplanetary space between the earth and the sun. At the present time, even this capability does not exist. Our present knowledge extends only to entry into the polar caps and entry to the synchronous altitude. Even that capability is uncertain.

To obtain the physical understanding, we need more analysis of the data that has been obtained which can address the topic (such as the Explorer 45 data base) and we need a properly instrumented satellite which covers the L range from 3 or 3.5 to at least 8. Pitch angle distribution data are required in order to determine the off-equatorial progression for solar proton flux injections.

## 2.2 The Requirements for Modeling and Prediction of Geomagnetic Storms

The importance of temporal variations in the trapped electron population in the slot and outer zone regions is qualitatively well known. In assessing the qualitative consequences of this on the modeling, forecasting, and user communities, it is important to be as restrictive as possible by focusing on user requirements in order to avoid unnecessary activity in the modeling and forecasting fields. It is now recognized that there are regions of the radiation zones in which the time scale of temporal variations is of the same order as the mission duration. As a result, missions which accumulate a significant flux contribution from these regions are not well served by the present generation of trapped electron models which contain time averaged representations of the particle population. In this role we focus on two sample missions in this category (noting that others exist) for which the requirements on the modeling and forecasting communities are radically different, partly on pragmatic grounds and partly as a result of differing user needs. The first is the normal mission analysis activity several years before launch, which is performed to predict the radiation environment. The second is EVA activity performed in association with man's presence at synchronous altitude.

In the first case, missions spending significant time in the L region of 3 to 5 may encounter an average flux over a 3-6 month period different by a factor of 4 or 5 from the flux 'predicted' by a model developed with the current technique of time averaging the data. The occurrence or lack of a storm event and the event size are the deciding factors. All that is positively known at the present time is that the mission has been shielded for a situation that is either unduly pessimistic or optimistic. The solution is to account for the effects of individual events. But to what extent? It is clear from the working group sessions that it is impractical to try to predict the occurrence of a single event in a given 3-6 month period with a lead time of years. It is questionable whether this can even be done meaningfully (i.e. quantitatively) with a lead time of hours (Higbie et al. 1979), hence for the long-term planners the question of prediction may be moot. This, however, does not mean that design should occur with respect to a worst case situation. As a realistic objective it should be possible to achieve two situations: 1) given rapid access to a ground based magnetic index (Dst?) to infer the peak storm flux to within a factor of 2-3 at any given L and the time of occurrence of the peak flux, and 2) by modeling the depletion phase of the storm to develop the time history of the flux at the satellite. This may be practical without an overwhelming modeling activity; however, does it have value to the mission planner? In some respects the answer may be no. The mission planner may require a model which can provide the



probability of occurrence of an event of given size in the mission period (not necessarily an easy task for modelers because of the number of observations - presently 7-10) and the event integrated flux resulting from the event. No per event prediction is involved per se. In some respects the answer to the question may be yes. If the above nominal radiation environment is assessed the nominal lifetime could be predictively updated in the event that a magnetic storm occurs; i.e., one could predict the demise of the mission. This implies a more rapid response from the modelers or the model users than presently exists. Further the value of this prediction to a military application is questionable unless a rapid launch capability exists.

In the second application, EVA activity at synchronous altitude, the role of prediction is a little clearer. The present generation of time-averaged models indicates that an EVA activity with  $.2 \text{ gm/cm}^2$  Al shielding (equivalent of the average space suit) reaches the currently valid mission radiation limit for a 90-day skin dose of 105 REM (about 81 rads Al) after only 1.35 hours of cumulative exposure. At the same time it is known that several orders of magnitude excursions of the flux occur around the average. The implications of a positive excursion are severe. The corollary is that the negative excursions, which persist for periods sufficient for individual EVA activity, have great potential. For this application ideally one needs to predict the onset and duration of a quiet period with sufficient lead time for scheduling EVA activity. Predicting the onset is practical and involves the definition of the depletion phase of a previously occurring storm. Prediction of duration may not be feasible explicitly. However, the working group on energetic particles has investigated the short-term prediction of the onset of a storm. If this is feasible, the appropriate question is the lead time of the prediction and the relationship to the time required for the EVA operator to reach shelter. Note that the prediction of the magnitude of the potential event is relatively unimportant since shelter must be sought irrespectively. However, the magnitude of the event would relate to the prediction of the onset of a quiet period. For large solar arrays several km in dimension the transit time to reach shelter may be several hours. Prediction of onset with this much lead time seems to be difficult. For large structures, transporters more heavily shielded than suits may be required to circumvent the problem.

The above applications relate to total dose. For rate dependent problems it may be possible to provide sufficient prediction time to influence operational scheduling in a meaningful way. As with the EVA activity, prediction of onset of a certain rate may not be possible with sufficient lead time for scheduling. However, a sufficient prediction of the time for the flux to return to a rate commensurate with equipment operation could be provided.

Almost all of the applications discussed above involve not only a predictive capability but also the ability within the modeling community to generate model fluxes on short notice. A new generation of model service is required in which models with some per-event flux capability are on-line on a computer with close operational links to both the mission project offices and the organizations generating the parameters to be used for prediction.

Again, almost all of the above applications require a knowledge of not only the flux at a given point in time but also at a given point in space. The present models are used in a local-time-averaged form by most users although a local time variation is available. The shielding community should investigate the use of the various L parameters including external magnetic field effects as a method of ordering particle data in the outer zone, eliminating the need for a local time variable. An additional

benefit would be a parameter simplification of the models in a spatial sense. It is noted that this general point has application to geosynchronous EVA activity since, although the satellite local time averages on a daily basis, the EVA activity of several hours duration may have a significant local time bias.

### 2.3 Energetic Solar Flare Proton Predictions

The integrated flux during the first four days after the August 1972 solar proton event exceeded the previous 11-year cycle integrated flux. In order to improve on the ability to predict the occurrence of these anomalously large (AL) events, it would be advantageous to investigate whether their occurrence correlates to any solar parameter or other observable solar phenomenon or event, and whether reliable precursor conditions can be established, of the order of hours-to-days, to be used for predictive purposes of AL events.

The calculation of event integrated annual totals of unattenuated, interplanetary solar flare proton fluxes at 1 AU for energies greater than 10, 30, and 60 MeV, from measurements obtained by the IMP series of satellites, demonstrated that these annual totals do not correlate well with sunspot numbers. It may be useful to investigate whether the actual occurrence of solar flare protons at 1 AU may be more reliably and accurately correlated to some solar parameter or phenomenon other than sunspot numbers.

Finally, penetration of solar flare particles into the magnetosphere needs to be considered in view of the following facts: a) local time variations of geomagnetic cutoff latitude are about 3-4 degrees, and b) storm effects can change these latitudes by up to about 2-4 degrees. These two variations are additive. Rigidity and geomagnetic shielding evaluations should take these variations into account. Preliminary estimates indicate that very significant differences in exposure values may result from this improvement.

### 2.4 Time Variations

For purposes of long-range mission planning (e.g. variable crew station periods, etc.), it may be useful to know whether a significant ( $\geq$  factor of 3) variation in the environment of critical duration (critical  $\approx$  30 to 60 day duration or periodicity) can be established, in order to bring crew schedules into phase with phenomena.

This is presently not possible since these variations, which may be large, appear to be stochastic. However, correlation with, for instance, fast solar streams may provide a mechanism for predictions. Decay is an important factor. Significant work is required in this area.

## 3. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS: PLASMA

Prepared by the Plasma Subgroup: H. B. Garrett, Chairman, J. Sroga, and A. Vampola

In the last few years an increasing emphasis has been placed on interaction between the low energy (10 eV - 100 keV) near-earth plasma environment and space systems. In parallel with this growth in emphasis has been a need for predictions of the



low energy plasma. As discussed in the review by Garrett (1979) these predictions are necessary because of the effects of a number of interactions. The Plasma Subgroup has attempted to summarize these interactions and identify the idealized parameters needed to understand and predict these interactions.

### 3.1 Basis of Requirements

The number of known interaction mechanisms between the low-energy plasma environment and space systems has steadily grown. Ranging from the effects of static charge buildup on satellites to the effects of argon beams on the plasmasphere, these interactions all have in common the necessity of knowing the distribution function of the ambient environment and how it evolves in time. The plasma subgroup has reduced the many interactions down to two primary types of effects. The primary effects and the specific aspects of the distribution function required are as follows:

1. Spacecraft charging - the process of charge buildup on spacecraft is not completely understood (specifically the arcing process and the plasma environment are not predictable at all). As a result, detailed knowledge of the distribution functions and composition as functions of time and position should be acquired until the critical parameters are identified. As discussed in Garrett (1979) and Garrett et al. (1979), however, the charging models at present require only the electron and ion currents and electron temperature. These quantities can be estimated either by statistical tables or by  $A_p$ . The detailed models of the environment will be required if any improvements are to be realized over these present models. It should be emphasized that in addition to the need to model the large scale plasma distributions within the magnetosphere, it is also essential to understand and to model the smaller scale characteristics of the hot plasmas (Johnson et al., 1977). For example, the charge distribution on a spacecraft could be greatly influenced by the highly anisotropic field-aligned electron and ion fluxes frequently observed as a plasma feature at geosynchronous altitude (McIlwain, 1975) and at lower altitudes over a wide range of L-shells and local times (Johnson et al., 1977). Also, the ionospheric components ( $O^+$  and  $H^+$ ) which are sometimes dominant in the hot plasmas can be highly structured spatially and/or temporarily (Geiss et al., 1979).

2. Contaminants - calculation of the deposition of ionized contaminants on spacecraft requires knowledge of the ambient particle distribution to determine not only the charging on satellite surfaces to which contaminants are attracted, but also the rate of ionization (Cauffman, 1973). The related problem of contaminant cloud propagation in the plasmasphere and magnetosphere (Chiu et al., 1979) requires detailed knowledge of the ambient population and, in order to estimate the evolution of the contaminants, the electric and magnetic fields - parameters dependent on geomagnetic activity.

The user groups can likewise be divided into two basic groups. The first group consists of the designers who are primarily interested in the specification of the space environment. Not only are they interested in specifying the geophysical environment before building a system in order to protect it from the above effects, but, as is often the case, also the specification of the environment for the purpose of investigating anomalies resulting from them. The second group is composed of forecasters - those who must predict when the anomalies may occur. This latter group is interested in two time periods. First, there is long-term prediction of 3 years or more for mission planning, and second, there are short-term (24 hour to 1 year) predictions. This last group is of most concern to this conference. Specific questions to be answered

are: a) How long will the plasma event last? b) What are the characteristics (composition, energy, and angular distributions, etc.) of the event? c) What effects will the event have on a particular system?

### 3.2 Requirements

From the preceding descriptions of interaction phenomena and of the user groups and required time scales, we believe it is evident the plasma distribution function and its evolution in time (or equivalently the electric and magnetic fields) are required in great detail - detail that is clearly not attainable now or in the near future. Thus simplification must be introduced. These are primarily real-time (either in-situ or from solar wind parameters) estimates of the environment and/or detailed statistical models of the plasma distribution function in terms of predictable parameters such as  $A_p$ . It is to this end this subgroup recommends current efforts be directed. Our state of knowledge in these areas is evaluated in the other working group reports.

## 4. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS: NEUTRAL ATMOSPHERE

Prepared by Neutral Atmosphere Subgroup: W. A. Vaughan, Chairman, R. Altrock, D. Hickman, D. Robbins, and J. Slowey

Orbital altitude total density and constituent number density variations are a direct function of the short- and long-term fluctuations in solar activity. These variations are due to the heating of the Earth's upper atmosphere by solar radiation and energetic particles. The importance of these variations is found in the requirement for orbital performance capabilities which will insure design lifetime, definition of orbital dynamics for nonspherical spacecraft, assessment of lifetime potential for spacecraft in orbit, and scientific experiments. Estimates of short- and long-term solar activity levels are critical inputs to these calculations.

### 4.1 Basis of Requirement

While there is a variety of users for neutral atmosphere models, we believe that their needs are reflected by the requirements of those using the models for orbital lifetime calculations. Therefore, this paper focuses on orbital lifetime prediction requirements.

A semi-analytical method is used in most spacecraft orbital lifetime prediction models to estimate the decay history and the lifetime of a near-Earth orbiting spacecraft perturbed by atmospheric drag. For most near-Earth orbits with small eccentricity, the perturbations due to other forces (i.e., solar-lunar gravity perturbations, solar radiation pressure, and electromagnetic effects) are overshadowed by the effects caused by uncertainties in the calculation of atmospheric drag. For this reason, efforts to incorporate additional perturbing forces are often unwarranted. The approach used to estimate the orbital decay usually adopts a combination of general and special perturbation techniques so that the analysis preserves sufficient rigor to insure accuracy and adequate numerical emphasis to include a rather sophisticated atmospheric density model in an efficient simulation. Basically, the procedure is to extend a system of ordinary differential equations for a set of well-defined mean orbital elements which describe the complete motion of a spacecraft about an oblate



Earth to include numerically the drag effect due to a rotating atmosphere. The program is designed to estimate, with reasonable accuracy, the orbital decay history and the orbital lifetimes efficiently and quickly. Figure 1 illustrates the principal inputs and components of a spacecraft orbital lifetime and decay prediction procedure.

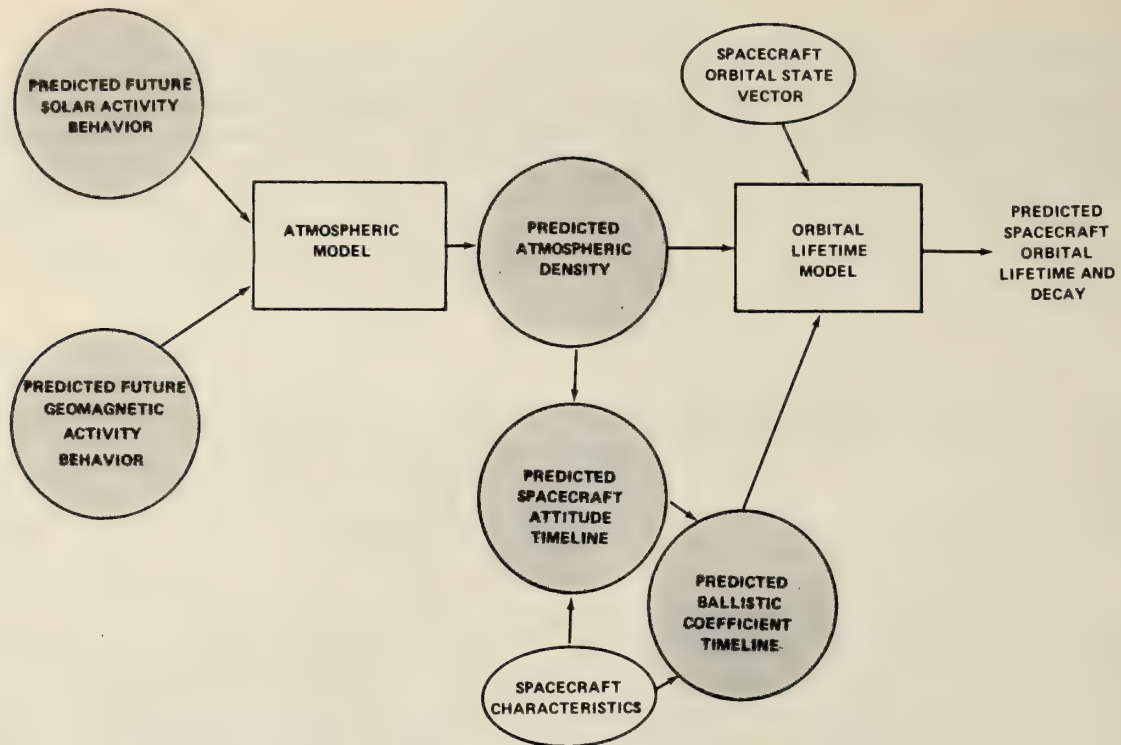


Fig. 1. Solar Predictions and Spacecraft Orbital Lifetime

A major difficulty in predicting orbital lifetimes arises because the future characteristics of the atmospheric density are not deterministically known. This makes it necessary to specify orbital lifetimes in a probabilistic manner. Comparisons of predicted spacecraft decay versus actual observed decay reveal deviations which can be attributed to an inadequate deterministic atmospheric model, noisy tracking data, or deviations in the stochastic variables associated with the lifetime prediction problem (i.e., ballistic coefficient, solar flux, and geomagnetic activity). The ballistic coefficient is a function of the spacecraft mass, drag coefficient, and effective cross section area. It is observed to vary with the spacecraft orientation and flight region. In addition, predictions of the solar flux and geomagnetic activity values, over either short or long periods of time, are available only in terms of statistical predictions with significant uncertainties.

Straus and Hickman (1979) describe the characteristics of several widely used atmospheric density models and have reviewed studies in which the predictions of these models have been compared with observational data. They also assess the relative advantages and limitations of the models in current use. They conclude that models produced prior to 1970 were developed from data bases with significant limitations. The overall accuracy of the recent models is summarized as being, on the average, as good as the measurements of atmospheric density and composition. This is evidenced

by the fact that some measurements have a mean deviation slightly higher while other measurements have a mean deviation slightly lower than the model calculations. We interpret this situation as being due to systematic instrumental inaccuracies. However, instantaneous measurements show a scatter of about a factor of two compared with the model calculations of atmospheric parameters.

The upper atmospheric density is strongly influenced by the changing levels of solar activity. This, in turn, directly affects the spacecraft drag and orbital lifetime. It is the ultraviolet solar radiation that heats and causes structural changes in the Earth's upper atmosphere. One component of this radiation relates to the active regions on the solar disk and varies from day to day, whereas the other component relates to the solar disk itself and varies more slowly with the 11-year solar cycle. Another influence on atmospheric density is due to energetic particles emitted by the sun as evidenced in the solar wind number density and velocity. These energetic particles are ultimately responsible for heating the Earth's upper atmosphere. The atmosphere reacts differently to each of these parameters and components (NASA, 1973 and Jacchia, 1977).

The 10.7 cm solar flux is generally used as a readily available index of solar ultraviolet radiation. It also consists of a disk component and an active region component. When the 10.7 cm flux increases, there is an increase in the upper atmosphere density. For a given increase in the disk component of the 10.7 cm flux, however, the density increases much more than for the same increase in the active region component. For all practical purposes the active region component is linearly related to the daily 10.7 cm flux and the disk component to the 10.7 cm flux averaged over a few solar rotations (e.g., six is used by Jacchia (1977)). The planetary geomagnetic index is generally used as a measure of the energetic particle heating. When the geomagnetic index varies we observe a density variation with a time lag of about 3 to 8 hours depending upon latitude.

An error, for example, of + 30 percent in the prediction of the maximum in the mean 10.7 cm flux during the ascending slope of the solar cycle, for a spacecraft at approximately 400 km altitude and having a nominal predicted lifetime of approximately 20 months, produces a decrease in the lifetime of approximately 30 percent. While this linear error relationship does not hold for all combinations of variations in solar activity, orbital altitudes, and ballistic factors, the example does illustrate the importance in the development of either a deterministic long-term solar activity prediction procedure or a statistical procedure with much closer confidence (error) bounds than now exists.

#### 4.2 Requirements For Solar Activity Predictions

There exists a critical need for more accurate predictions of short- and long-term solar activity to use in atmospheric density models. This will be required not only for the monitoring and accurate estimation of the orbital lifetime and decay for the large numbers of spacecraft and "junk" now in orbit but for the more economical and efficient estimation of future spacecraft missions, especially in the near-Earth orbital environment. The expected future state of the upper atmosphere plays an important part in the decisions associated with spacecraft missions. The critical dependence on solar activity predictions can easily be seen by looking at the flow chart shown in Figure 1.

That a neutral atmosphere model must be accurate is obvious. Less obvious is



how the accuracy is to be characterized for a particular use. A perfect description of the atmosphere is the ideal, but some kinds of inaccuracies are unimportant for specific applications. A spacecraft designer who is required to provide an orbital altitude capability for a two year lifetime may be quite satisfied with a model output having the correct mean value when averaged over a month's time span even if the errors averaged over a single orbit are quite large. On the other hand, a tracking station operator or mission monitor concerned with spacecraft decay within a few days or weeks may attach no importance at all to the long-term mean behavior of the atmosphere if the model output can accurately predict the short-term variations.

Current neutral atmosphere models employ 10.7 cm flux and geomagnetic index ( $A_p$  or  $k_p$ ) as heating indicators for model inputs. The solar-terrestrial prediction requirements to meet most user needs are given in Table 1.

Table 1. Solar-Terrestrial prediction requirements.

Approximate Prediction Period	Critical Parameter	Accuracy	Resolution	Frequency of Update
Very Short-Term (< 1 mo)	$F_{10.7}$	5%	6 solar rotations	Weekly
	$F_{10.7}$	5%	Day	Daily
	$A_p$	5%	Day	Daily
Short-Term (1-2 mo)	$F_{10.7}$	5%	6 solar rotations	Monthly
	$F_{10.7}$	5%	Day	Weekly
	$A_p$	5%	Week	Weekly
Long-Term ( $\geq 3$ mo)	$F_{10.7}$	10%	6 solar rotations	Quarterly
	$F_{10.7}$	10%	Week	Monthly
	$A_p$	10%	Quarter	Quarterly
Long-term ( $\geq 1$ yr)	$F_{10.7}$	10%	6 solar rotations	Yearly
	$A_p$	10%	Year	Yearly
Very Long Term (1-2 cycles)*	$F_{10.7}$	10%	6 solar rotations	Start & Peak of Cycle
	$A_p$	10%	Year	Start & Peak of Cycle

\* Maximum and Minimum Values of Parameter Plus Dates of Maximum and Minimum Occurrence

Why is it that neutral atmosphere models do not do a better job of predicting orbital altitude density and number density of the constituents? A very major part of the answer lies in the inaccurate predictions of short- and long-term solar activity and geomagnetic index values used as inputs in the models. Part of the answer also lies in the selection of parameters to characterize conditions of the orbital neutral atmosphere. For example, the amount of EUV heating is represented by the 10.7 cm solar flux. The 10.7 cm flux cannot significantly heat the atmosphere. However, it has been established that there is a reasonable correlation between the EUV flux and the 10.7 cm flux. This correlation is only approximate and is not adequate for models yielding high accuracy. However, the 10.7 cm flux is measured regularly and hence is

readily available. A similar situation exists with respect to the energetic particle heating which is often measured by the geomagnetic index. This heating is related to disturbances in the earth's magnetic field, but the relationship also may not be adequate for high accuracy models.

The future development of neutral thermospheric models capable of providing more accurate predictions depends critically on replacing such indices as the 10.7 cm flux and  $A_p$  with new parameters based more closely on the physical quantities which affect the upper atmosphere. Such new parameters must directly characterize the UV/EUV flux which is the primary source of energy into the thermosphere and must indicate the thermospheric heat input into high latitude regions caused by particle precipitation. The means to develop these new parameters is available, but careful planning of future missions is required if acquisition of adequate information is to be ensured.

The exoatmospheric solar UV/EUV flux in the range from  $300\text{\AA}$  to  $2000\text{\AA}$  should be monitored to determine directly the major thermospheric heating. The precipitating particle flux (in an energy range from a few hundred eV to several keV) can be measured from low altitude polar orbiting satellites as is being done on TIROS-N and DMSP. This will allow a quantitative relationship to be established relating atmospheric response to high latitude heat sources. Finally, in order to understand the precipitation mechanisms, solar wind density and velocity should be monitored by sensors sunward of the magnetosphere. The response of the atmosphere should be monitored by density- and composition-measuring instruments having good resolution in both time and space. It should be emphasized that if unambiguous interpretation is to be made of the results of these measurements, all parameters should be measured simultaneously.

The measurements which must be made in order to ensure significant improvement in neutral density and composition models for better prediction of satellite orbit evolution are summarized in Table 2 and it is strongly recommended by this Neutral Atmosphere Subgroup that future mission plans be made to ensure simultaneously obtaining all of these significant parameters.

Table 2. Measurements required for development of significantly improved neutral density models

Parameter to be measured	Parameter Range	Location of Measurement	Purpose
Solar UV/EUV Flux	$300\text{\AA}-2000\text{\AA}$	Above $\approx 200$ km	Measure primary source of energy into the atmosphere
Precipitating particle flux	Few hundred eV to few keV	Low earth orbit (Polar)	Measure high latitude heat source during magnetic storms
Solar wind density, velocity	—	Sunward of magnetosphere	Understand precipitation mechanisms
Atmospheric density composition	—	$\geq 130-800$ km	Determine atmospheric response to energy inputs

Note: For unambiguous understanding of observed phenomena, all quantities should be measured simultaneously.



### 4.3 Concluding Remarks

The prediction requirements given in Table 1 for solar 10.7 cm flux and planetary geomagnetic index parameters are based on first hand knowledge and experiences of the neutral atmosphere subgroup members as users, consultants to users, atmospheric model developers, scientists, experimenters, and predictors of solar activity to meet their own user needs. For a service or research organization concerned with or interested in trying to meet these requirements a logical question to ask concerns how serious are the users. Will the requirements simply fade away when a specific user is challenged? Will the requirements be radically relaxed or disappear when one of these organizations says to a specific user - give us the funds and we'll embark on a program to meet these requirements? The answer is probably yes to one or both questions but the degree depends upon how critical the requirement is to the particular project for which the user needs the predictions, the risks he's willing to take, the project's schedule, trade-offs on these requirements versus other project requirements, his confidence the proposal will produce results he can apply for the benefit of his project, and the costs.

Therefore, these requirements are not rigorous for all users but depend upon many factors unique to each user and his immediate project needs. For example, major decisions are made on spacecraft orbital altitudes which depend on the current inaccurate solar activity predictions. In some cases this results in less than desirable orbital altitudes relative to the scientific experimenter's requirements, higher mission success risks, provisions for much longer lifetimes than needed due to the large error bounds on the orbital altitudes estimates for the mission, and added costs for spacecraft instrumentation, operational capabilities and decay monitoring to say nothing of the embarrassment which results when an expected small risk spacecraft lifetime mission design is significantly different than expected prior to launch of the spacecraft. Thus these requirements should be taken as serious candidates for technology and scientific research program sponsorship by responsible service and research organizations.

## 5. USER REQUIREMENTS OF SOLAR-TERRESTRIAL PREDICTIONS FOR SPACECRAFT APPLICATIONS: RECOMMENDATIONS

The following constitutes a summary of the recommendations contained in the previous sections. It is separated by particle type to conform with the previous discussions. A warning: For an operational program, the cost of obtaining and/or using predictive techniques must be less than the cost of going to another hardware or mission design which avoids the hazard.

### 5.1 Energetic Particles

1. A requirement exists for a predictive technique which relates some solar or geophysical parameter (such as solar wind speed or  $D_{st}$ ) to acceleration of electrons to high energies in the outer magnetosphere. The required output is: a) Spatial Distribution, b) Energy Spectra, and c) Flux Intensities.
2. Given an input distribution (spatial, energy spectrum, flux intensity) of energetic electrons in the outer magnetosphere, a model is needed which will detail the evolution of the distribution.

3. Manned applications require a short term prediction of magnetic storms and substorms. Particle acceleration and plasma injection are both major concerns.
4. A 'Disturbance Model' for prediction of flux enhancements due to solar or magnetospheric activity is required. It should have two forms: a typical storm, and a macro storm.
5. A requirement for synoptic measurements of solar wind parameters to predict intermediate term (30-60 days) averages of fluxes at synchronous altitudes exists.
6. A model which calculates the hardening of electron energy spectra in response to rapid diffusion caused by field-line loading is desired.
7. Current methods take several days to calculate the environment and dose for a new orbit. This must be shortened.
8. The long lead time (typically several years minimum) to get new data into a data base for modeling purposes must be shortened.

## 5.2 Solar Flare Protons

1. A predictive technique must be developed which will warn of anomalously large events (e.g., August 1972) hours or days in advance. Identification of precursors is the most probable method.
2. An accurate prediction of solar proton events based on solar parameters should be developed. It should give order-of-magnitude or better definition of the intensities and energy spectra. Timing is of concern.
3. Given an event on the sun, predict the evolution of an event from solar or interplanetary parameters.
4. Models of solar proton entry into the magnetosphere which include local-time and magnetic storm effects in rigidity calculations should be developed.
5. Determine the solar parameters which correlate well with the annual integrated fluences of unattenuated interplanetary protons with energies above 10, 30, and 60 MeV, since sunspot numbers do not.

## 5.3 Plasma

1. A statistical model of the ambient low-energy plasma is required, (i.e., percentages of time certain conditions will be encountered).
2. A three-dimensional model of magnetospheric plasma distribution is needed.
3. Verification of plasma drift theories and understanding of the processes which energize ionospheric ions and inject them into the magnetosphere is required in order to detail the evolution of the hot plasma distribution.
4. Expanded real-time measurements (in longitude and altitude) are desired for predictions for operational programs.



5. A better understanding of substorms and their plasma effects is required in order to ultimately produce predictive techniques.
6. Measurements of interplanetary particles and fields are desired to assist research into predictive techniques.

#### 5.4 Neutral Atmosphere

1. Continuing synoptic observations of the Ottawa 10.7 cm flux index and the  $A_p$  geo magnetic index are required. (Details of the frequency and accuracy of updating are given in Table 1.
2. In order to develop better predictive models of the atmosphere, the following parameters should be measured simultaneously: a) Solar UV/EUV above 200 km; b) Precipitating particle flux; c) Solar wind density and velocity sunward of the magnetosphere; d) Atmospheric density and composition between 130 and 800 km.

#### 6. MINORITY REPORT A. L. Vampola

The following recommendation was considered by the working group but was not formally accepted. It is included here because of the enthusiasm with which it was received by some participants in some of the other groups and because an essentially identical recommendation was offered for the group's consideration by a modeler.

Recommendation: That a centralized catalog of 'State of the Art' models and predictive techniques in solar-terrestrial phenomena be established and maintained by one of the prediction or data archival service agencies (such as SESC, WDC-A, or AFGWC).

The rationale behind this recommendation is that the 'Corporate Memory' in spacecraft engineering seems to reside in old proposals and contract specifications. The normal evolution of an engineering career is to become part of management after a number of years of design engineering. Technologies become obsolete and so do models and predictive techniques. A new engineer is usually forced to rely on documents inherited from the previous occupant of his 'slot' who has moved on and is no longer keeping track of the minutiae of his previous position. As a result, it is not uncommon to see models or predictive techniques written into contracts ten or more years after they have been supplanted.

By having a central catalog for these techniques/models, the latest version would be readily identified by potential users. It would be incumbent upon the modelers, themselves, to ensure that the latest versions of their works were listed in the catalog. Distribution of computer programs or large data bases required by some models would continue as at present: through direct contact between the user and the modeler. In the case of substantial changes in modeling or predictive techniques, some international group (similar to the present workshop or established single interest scientific bodies) could make the decision to abandon one model or technique and supplant it with another in the catalog.

## 7. REFERENCES

- Atwell, William (1979): Dosimetry in the manned space program. These proceedings.
- Baker et al. (1979): The use of  $\geq 30$  keV electrons anisotropies at  $6.6 R_e$  to predict magnetospheric substorms. These proceedings.
- Cauffman, D. P. (1973): Ionization and attraction of neutral molecules to a charged spacecraft. SAMSO TR-73-263.
- Chiu, Y. T., J. M. Cornwall, J. G. Luhmann, and Michael Schulz (1979): Argon-Ion contamination of the plasmasphere. SSL-79(7824)-1, The Aerospace Corporation, El Segundo, CA.
- COSPAR (1972): COSPAR International Reference Atmosphere, 1972, Akademie-Verlag, Berlin.
- Garrett, H. B. (1979): Low energy magnetospheric plasma interactions with space systems - the role of predictions. This Volume.
- Garrett, H. B., A. G. Rubin, and C. P. Pike (1979): Prediction of spacecraft potentials at geosynchronous orbit. These proceedings.
- Geiss, J., H. Balsiger, P. Eberhardt, H. P. Walker, L. Weber, D. T. Young, and H. Rosenbauer (1979): Dynamics of magnetospheric ions as observed by the GEOS mass spectrometer. Proceedings of the 13th ESLAB Symposium, Space Sci. Rev., in press.
- Hedin et al. (1977a): A global thermospheric model based on mass spectrometer and incoherent scatter data, MSIS 1, N2 density and temperatures. J. Geophys. Res., **82**, 2139.
- Hedin et al. (1977b): A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 2, composition. J. Geophys. Res. **82**, 2148.
- Higbie, P. R., D. N. Baker, R. D. Belian, and E. W. Hones, Jr. (1979): Evolution of substorm and quiet-time electron anisotropies ( $30 \leq E_e \leq 300$  keV). These proceedings.
- Jacchia, L. G. (1977): Thermospheric temperature, density and composition: New models. Special Report 375, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.
- Johnson et al. (1979): Solar induced variations of environments affecting manned space flight and spacecraft operations. These proceedings.
- Johnson, R. G., R. D. Sharp, and E. G. Shelley (1977): Composition of the hot plasma near geosynchronous altitude. In Proceedings of the Spacecraft Charging Technology Conference, NASA TMX-73537, C. P. Pike and R. R. Lovell ed., p. 53.



- Kolasinski, W. A., J. B. Blake, W. E. Price, and E. Smith (1979): Simulation of cosmic ray induced soft errors and latchup in integrated-circuit computer memories. To be presented at the 1979 IEEE Nuclear and Space Radiation Effects Conference.
- McIlwain, C. E. (1975): Auroral electron beams near the magnetic equator. Physics of the Hot Plasma in the Magnetosphere, B. Hultqvist and L. Stenflo, ed., Plenum Pub. Corp., N. Y., p. 45.
- NASA Vehicle Design Criteria (1973): Models of earth's atmosphere (90 to 2500 km). NASA SP-8021, National Aeronautics and Space Administration, Washington, D.C.
- Sivo, L. L., M. Brettschneider, W. Price, and P. Pentecost (1979): Cosmic ray induced soft errors in static MOS memory cells. To be presented at the 1979 IEEE Nuclear and Space Radiation Effects Conference.
- Stassinopoulos, E. G. (1979): The terrestrial radiation environment and EVAs: Prediction requirements, model improvements and warning systems. These proceedings.
- Straus, J. M., and D. R. Hickman (1979): Predictability of upper atmospheric density and composition. This Volume.
- Vette, J. I., M. J. Teague, D. M. Sawyer, and K. W. Chan (1979): Modeling the earth's radiation belts. This Volume.

# MODELING THE EARTH'S RADIATION BELTS

(A Review of Quantitative Data Based Electron and Proton Models)

J. I. Vette, M. J. Teague, D. M. Sawyer, and K. W. Chan  
National Space Science Data Center  
NASA/Goddard Space Flight Center  
Greenbelt, Maryland 20771

The evolution of quantitative models of the trapped radiation belts is traced to show how the knowledge of the various features has developed, or been clarified, by performing the required analysis and synthesis. The Starfish electron injection introduced problems in the time behavior of the inner zone, but this residue decayed away, and a good model of this depletion now exists. The outer zone electrons were handled statistically by a log normal distribution such that above 5 Earth radii there are no long term changes over the solar cycle. The transition region between the two zones presents the most difficulty, therefore the behavior of individual substorms as well as long term changes must be studied. The latest corrections to the electron environment based on new data are outlined. The proton models have evolved to the point where the solar cycle effect at low altitudes is included. Trends for new models are discussed; the feasibility of predicting substorm injections and solar wind high-speed streams make the modeling of individual events a topical activity.

## 1. INTRODUCTION

The purpose of this paper is to review the current status of the quantitative models of the energetic particles in the Earth's radiation belts and to discuss the trends expected in future modeling efforts. From the first observations of the magnetosphere beginning with Explorer 1, flux maps of the trapped electrons or protons were made from the results of individual satellite experiments and the closure of the contours was estimated to obtain an overall picture. To differentiate between models and this type of analysis, or presentation, we consider a model to consist of the following: (1) a synthesis of the results from experiments on a number of satellites; (2) a statement or algorithm on the time behavior, so that the product can have other than historical use; and (3) a quantitative description of the flux and spectral distribution that can readily be used by others.



The first such models were constructed by Wilmot Hess following the Starfish detonation in July 1962, when a number of satellites were affected by the extremely large flux levels of energetic electrons. These models were documented through letters to a group of users. Following this series, an effort was begun by James I. Vette, with NASA and USAF support, beginning in early 1964 at the Aerospace Corporation and carried out since 1967 at the National Space Science Data Center (NSSDC) with the help of a number of colleagues. In this paper the important features that have gone into the models will be presented. This will show the evolution of this work and provide the framework to discuss other characteristics that may be possible to incorporate in future efforts. All of the models completed by our group were extensively documented, so a detailed presentation of them here is not warranted.

Flux mapping based on individual results continue to be presented by some investigators, and such efforts provide a useful input for the production and improvement of the NSSDC models. However, the range of validity of such maps should be clearly understood by users; the failure to do so has caused confusion in some cases. Because all the data that were presented in these individual maps either were or will be incorporated into our models, none of them will be discussed here explicitly.

## 2. THE EVOLUTION OF ELECTRON MODELS

The first electron model, AE1, was derived from 10 experiments on 9 different satellites covering the period October 1962 through November 1963 (Vette, 1966). The designation of these models included the letter A for Aerospace to distinguish them from the E1, 2, ..., 8, and P grids of Hess. The model covered the regions up to  $L = 3$ , where  $L$  is the McIlwain parameter, and energies between 0.3 and 5 MeV with extrapolation to 7 MeV. Using available energy calibrations for the various detectors, the data were handled, for the first time, in a consistent manner to obtain the fluxes from the observed counting rates by using the model spectrum. This was accomplished in an iterative fashion because a spectrum had to be chosen to begin the process. Because of a number of factors, the small dependence of the spectral parameters on the magnetic field intensity,  $B$ , was not able to be represented. The decay of the Starfish residue was the dominant effect, and decay times were assumed to be energy independent. Although this allowed the determination of flux values for future time periods, the level of the natural inner zone electrons could not be provided at that time. Hess had used a similar procedure with his last electron model, E8. A two-dimensional grid in  $B$  and  $L$ , for the spatial distribution, and a similar one in  $E$ , the energy, and  $L$  for the spectrum were provided on punched cards. In addition, as has been the practice with the documentation of our models, extensive graphical and tabular displays of the fluxes were given.

In the production of AE2, data from 15 instruments on 9 satellites that spanned the period October 1962 through October 1964 were utilized (Vette *et al.*, 1966). The energy range was extended downward to 40 keV and the  $L$  range extended outward to 6 to include the peak of the outer zone. An exponential spectrum above 2.5 MeV was assumed with no  $B$  dependence. To handle the large

dynamic time variations that were observed above  $L = 2$ , a crude average of the logarithm of the flux was used as the model value. Decay of the inner zone was handled in the same manner as in AE1; the same card deck format was also employed to distribute this model. In addition, a crude model was projected for 1968 estimating Starfish decay from August 1964, the epoch of AE2, and outer zone solar cycle behavior. (This latter effect was grossly overestimated at the time.)

Because of the extreme interest in the geostationary orbits, the next effort was a model, AE3, for the single  $L$  value, 6.6. It is interesting to note that this was completed prior to the launch of ATS 1, the first synchronous satellite to carry radiation detectors for electrons and protons. Several significant improvements were incorporated into the analysis at this time. It was necessary to employ a third spatial coordinate because the distorted magnetic field caused by boundary, tail, and ring currents could not be represented by the conventional B-L system computed from the internal geomagnetic field. Local time,  $\phi$ , was chosen along with the conventional B and L parameters as a geometric coordinate system that would blend into the physically meaningful B-L system below  $L = 4$ . The data from 12 instruments on 6 elliptically orbiting satellites covering the time period August 1959 through October 1965 were used to produce AE3 (*Vette and Lucero*, 1967). Besides determining the  $\phi$  variation of the flux above  $L = 5$ , a detailed treatment of the time behavior was accomplished. It was shown that a reasonable approximation to the first two moments of the time distribution observed over 6 months or more was a log normal distribution, and assumed that each data point was uncorrelated, which is certainly not true, but it did offer a practical way to determine the fraction of the time that the flux exceeded a certain level. A quantitative determination of the average flux levels over each data set was obtained. Although the differences between data sets were typically a factor of 3, it was possible to make a statement that there did not seem to be any long term changes in these flux averages, particularly any caused by solar cycle effects. This was borne out, as will be discussed later. The model was given in analytical form as the product of four functions: (1) an exponential in E, (2) a power law in E with a small exponent dependent on  $\phi$ , (3) a power law in B, and (4) a tabular function of  $\phi$ . Tables of the fluxes as functions of the four variables were provided. The standard deviation of the log normal distribution was given by

$$\sigma = 0.62 E^{0.2} \quad (1)$$

Statistical tables were computed giving the fraction of the time that the flux would exceed a given level for energies from 0.01 to 6 MeV. These were computed only at the equator with a local time-averaged model.

Following this effort, a study of the entire outer zone was done by applying the techniques developed for AE3. Twenty-three data sets from 11 satellites covering the period from August 1959 through February 1968 were analyzed to construct AE-4 (*Singley and Vette*, 1972a). Four important data sets incorporated into this model were those of *Paulikas and Blake* (1971) on ATS 1, which provided the most definitive picture on the local time variation of outer zone fluxes, and extended the energy range at these altitudes to 2 MeV. Unfortunately, low altitude data were not readily available for the type of processing required. The form of the model was similar to AE3 with an L



dependence added, and the  $\phi$  dependence was simplified using ATS 1 results to

$$\log J \propto C(E, L) \cos \left( \frac{\pi\phi - 11}{12} \right) \quad (2)$$

with  $\phi$  in hours and  $C(E, L \leq 5) = 0$ . To provide a low-altitude cutoff, the B variation was given by

$$\left( \frac{B}{B_0} \right)^{-m(L)} \left( \frac{B_c - B}{B_c - B_0} \right)^{m(L) + 1/2} \quad (3)$$

employing the flux pair theorem of *Roberts* (1965).  $B_0$  is the magnetic field at the equator, and  $B_c$  is the cutoff value determined such that the maximum altitude reached by  $B_c$  was 100 km. The model covered the L range from 2.8 to 11, and the energy range 0.04 to 5 MeV. The model was distributed as a computer program that included AE-5, the inner zone companion to AE-4. Both a solar minimum and a solar maximum version were produced. Statistical tables were also provided (*Singley and Vette*, 1972b).

The most important results from this modeling effort were the demonstration that the  $\phi$  variation was unimportant below  $L = 5$ , and that the solar cycle effects occur only below  $L = 5$  all the way into the inner zone around  $L = 1.8$ . This is illustrated best in Figure 1, where data from the identical instruments of Winckler on OGO 1 and OGO 3, which were intercompared in orbit, were processed to obtain the results. The effect of the Starfish fluxes was still present at the higher energies. It is apparent that there is very little change above  $L = 5$ , but the absence of a solar cycle effect at the geostationary altitudes is most effectively demonstrated by the recent work of *Paulikas and Blake* (1978), who intercompared their instruments on ATS 6 with those of *McIlwain* on ATS 5 and presented the time history, with some gaps, from the launch of ATS 1 in December 1966 through July 1977. An example of this work is shown in Figure 2. The values obtained with AE-4 are at the appropriate B, L point in order to correspond with the longitude of the satellite. The model is high at 0.7-MeV energy and low at 3.9 MeV. For the two lowest energies the long term averages are seen to be relatively constant. The seemingly stochastic behavior of the highest energy was shown by *Paulikas and Blake* (1978) to bear some correlation with the occurrences of high-speed solar wind streams. These streams also affect the lower energy fluxes, but with a faster decay time the long term averages come out about the same.

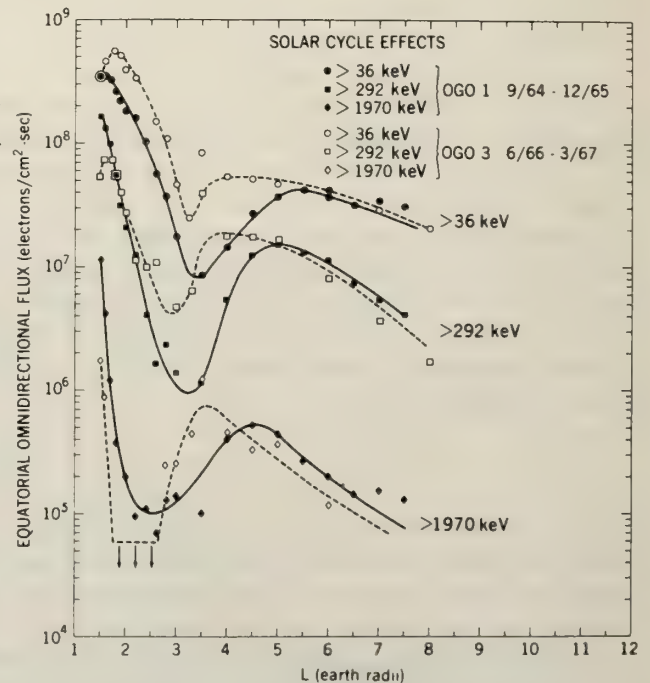


Figure 1. Solar cycle effects using results from the OGO 1/OGO 3 electron spectrometers. The data were time averaged over the periods shown and converted to integral spectrum and to omnidirectional equatorial flux.

The AE-5 model was completed concurrently with AE-4 and covered the L range 1.2 to 2.8 (*Teague and Vette, 1972*). The data analyzed for this model came from six instruments on five satellites that covered the period September 1964 through December 1967. The model adopted in this region assumed that the total electron flux consisted of four components: (1) a quiet-time background flux at solar minimum, (2) a quiet time flux that varied with the solar cycle, (3) an impulsive flux from a magnetic substorm that resulted in enhanced flux levels, and (4) the residue from the Starfish injection.

The fact that Starfish electrons at all energies had decayed above  $L = 2$  by September 1964, and lower energies to  $L = 1.4$  by December 1966, allowed the determination of (1) and (2) above.

The Starfish decay model of *Stassinopoulos and Verzariu (1971)*, along with further analysis of the data, permitted the determination of (4). By examining the large impulsive increases between June 1966 and December 1967, the time-averaged flux values in the L region from 1.8 to 2.8 were obtained. No substorm increases were seen below  $L = 1.8$  in this time period.

The significant advances in the modeling efforts as a result of the analysis for AE-5 were: the B dependence of the energy spectrum was taken into account with a functional form,  $E \exp[E/E_0(B, L, T)]$ , where T represented the time since the Starfish injection; the solar cycle variation between  $L = 1.4$  and 2.4 from a T of 6 to 37 months was ascertained for energies below 300 keV; and the ratio of the magnetic substorm impulsively-injected flux to the quiet time flux at epoch October 1967 was obtained (the peak of the sunspot number for the 20th cycle was nearly reached by this time).

The separation of the Starfish and solar cycle components is shown at two L values in Figure 3, and the average substorm effects are given in Figure 4 as a function of E and L.

Using the concepts developed for AE-5, *Teague and Stassinopoulos (1972)* were able to improve the Starfish decay model of *Stassinopoulos and Verzariu (1971)* and to determine the times that the Starfish residue disappeared below

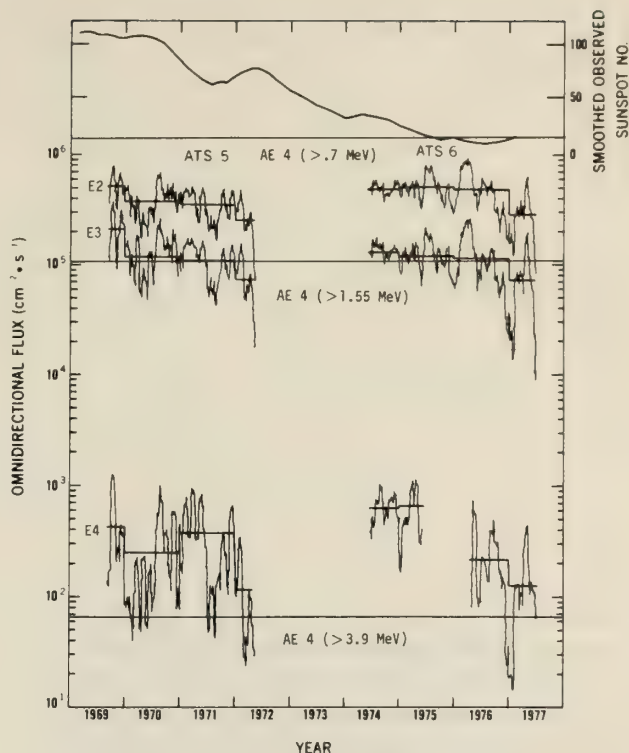


Figure 2. Twenty-seven day running flux averages from ATS 5 and ATS 6 compared with sunspot number. The short solid bars are the average values over the portion of the calendar year. The horizontal lines extending across the graph are the fluxes obtained from the AE-4 model.



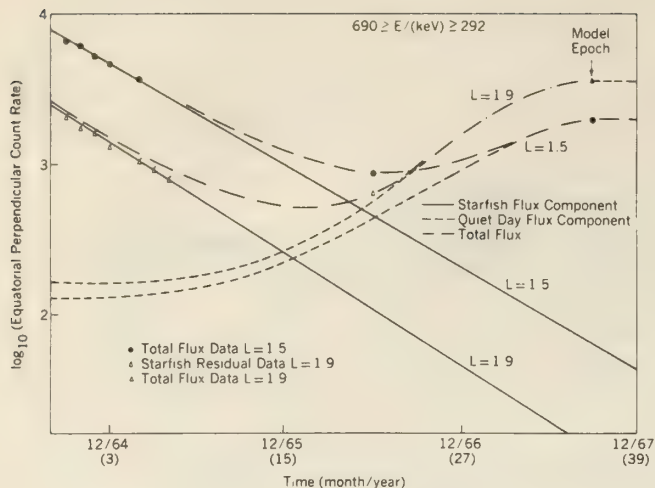


Figure 3. Separation of Starfish decaying and solar cycle-varying electron flux components. The numbers in parentheses on the abscissa scale are the months from September 1964 (the assumed solar minimum for this work).

data, incorporating the solar cycle model, and using the Starfish decay model shown in Figures 5 and 6, *Teague and Vette* (1974) produced an inner zone electron model for solar minimum conditions without any Starfish electrons. This model was denoted AE-5 1975 projected. Using the same techniques, AE 6 was produced to describe the 1980 solar maximum conditions (*Teague et al.*, 1976). All of these inner zone models were incorporated with the appropriate version of AE-4 into a computer program first described by *Teague et al.* (1972), which was used to distribute these environments to the users.

A discussion of the new model AE-7 presently under construction will be given in Section 4. This model will cover the whole trapping region and will utilize the latest data available in order to make improvements.

the natural background by using the data from eight satellites. The decay contours for this model are shown in Figure 5. The B dependence in the decay time,  $\tau$ , was obtained for  $L < 1.4$ . The cutoff time contours for the disappearance of Starfish decaying electrons are given in Figure 6, which shows that by mid-1970 all of these electrons had become nearly imperceptible.

By making some corrections to AE-5 based on a further analysis of

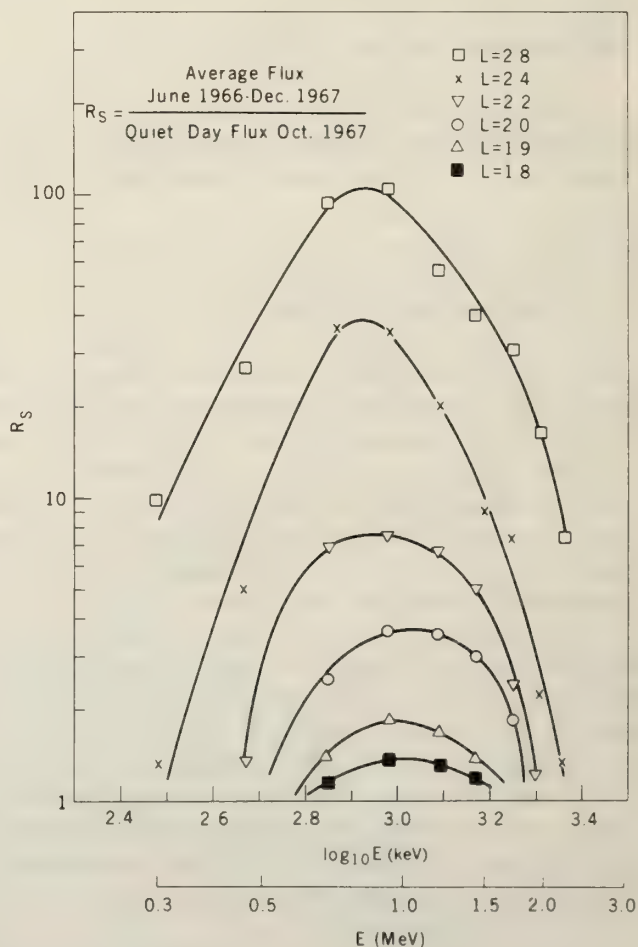


Figure 4. Ratio of time-averaged electron flux over 18 months to the quiet time flux in October 1967. This illustrates the effective enhancements caused by large magnetic substorms.

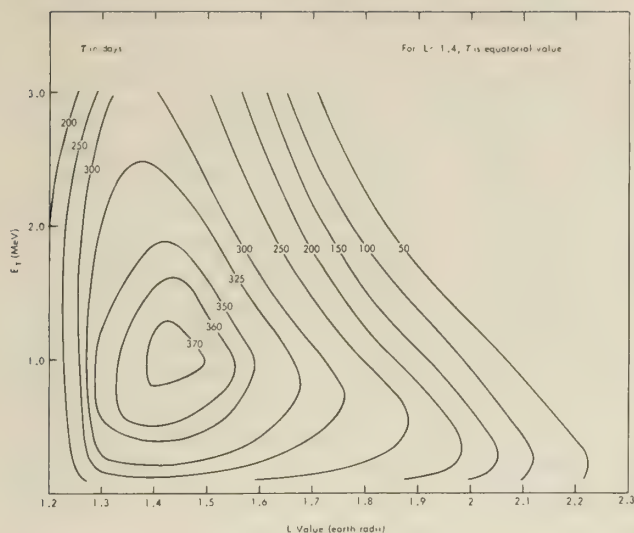


Figure 5. Constant decay time contours for Starfish electrons.

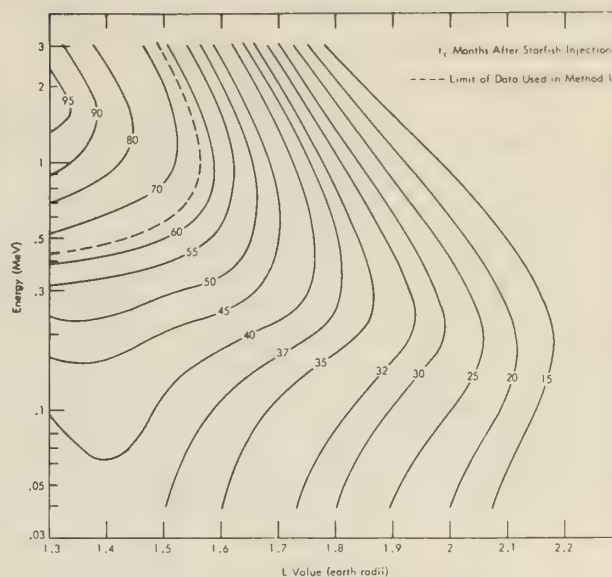


Figure 6. Constant cutoff time contours for Starfish electrons.

### 3. THE EVOLUTION OF PROTON MODELS

While the evolution of electron models was driven by the need to account for the many time variations in a more precise manner, proton models have changed mainly to improve the description of the energy dependence. The first four models, AP1, 2, 3, and 4 covered the energy ranges  $> 30$  to  $> 50$ ,  $> 15$  to  $> 30$ ,  $> 50$ , and  $> 4$  to  $> 15$  MeV, respectively. Each had an exponential energy spectrum dependent on B and L. These were constructed separately using the data from 15 instruments on 10 satellites and 1 rocket flight, covering the time span July 1958 through September 1963 (Vette, 1966). They were distributed in the same card format as AE1 and AE2. All were static models, although solar cycle and magnetic substorm effects of factors generally  $< 2$  were discussed. One of the shortcomings of these models was the failure to provide a smooth joining at the energy boundaries over all B and L.

King (1967) extended the energy range to  $> 0.01$  MeV with the production of AP5, again employing an exponential spectrum dependent upon B and L. The data from nine instruments flown on six satellites covering the period August 1961 through April 1965 were used in the analysis. By including new measurements from five detectors on three satellites and using a power law spectrum, Lavine and Vette (1969) were able to cover the energy range  $> 4$  to  $> 30$  MeV satisfactorily in producing AP6. Again it was determined that time variations were not significant, relative to the disagreement in the various data sets. With the examination of new data covering the period August 1964 through July 1966 and using the older data for a total of 21 detectors on 12 satellites, a more comprehensive model above 50 MeV (AP7) was constructed by Lavine and Vette (1970). An exponential spectrum was still found to be the best for this energy region. The spread between the model and the various



data sets was always less than a factor of 3; an error of a factor of 2 or less was more typical.

Finally, by combining the results of 94 different energy channels on various instruments flown on 24 different satellites, *Sawyer and Vette* (1976) produced AP-8, covering the entire energy range from 0.1 to 400 MeV. With the introduction of the AZUR data in 1969-70 and the OV3-3 data in 1966, it was possible to represent solar cycle variations below 1000 km with  $L < 3$ . This was accomplished by producing both AP-8 MIN and AP-8 MAX, which differ only in the region just mentioned. An example of the effect is shown in Figure 7; at high B values the effect can reach a factor of 10 or more. The volume of space where this effect occurs is small, but may be important for future long term Space Shuttle flights.

A good summary of the other temporal variations that were observed in proton fluxes is given in *Sawyer and Vette* (1976), but it is too detailed to present here. The AP-8 models are available in a computer program similar to that used to distribute the latest electron models. Because the block data statement is too large for some computers, a smaller version is also available that provides values that are accurate to better than a factor of 2 of the full model. Because there are no longer separate energy regimes, this model eliminated the energy boundary problems associated with the earlier models.

#### 4. FUTURE TRENDS IN MODELING

The future trends will be discussed in three different categories. The first involves the improvements we hope to realize, utilizing new data that have become available and using methods suggested by a better understanding of the important magnetospheric processes. The second category involves the utilization of data on presently operating satellites. However, it may be 5 or 6 years before such data are available, if the past can serve as a guide. Finally, we will present some remarks based on recent observations that suggest the possibility of making short term predictions of enhanced fluxes. Besides utilizing these predictions for operational missions, they should be

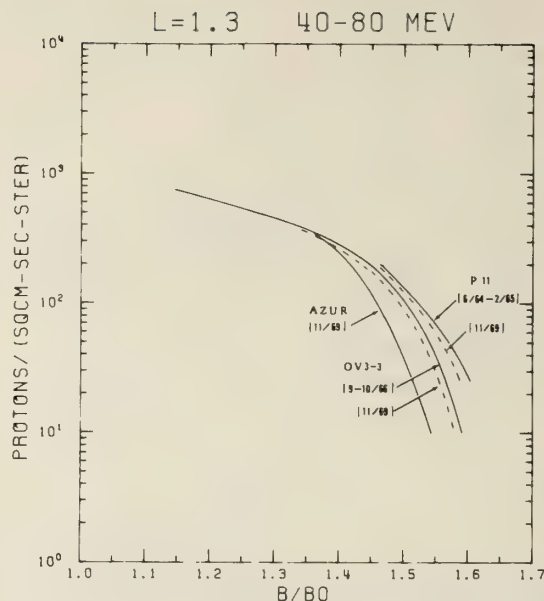


Figure 7. Solar cycle effect on energetic protons at low altitude. The time period for each of the three data sets is shown. The dotted lines show the expected position of the AZUR data for the time corresponding to the midpoints of the P11 and OV3-3 data, according to the theory of *Dragt* (1971).

used to improve coordinated data acquisition, which would result in increased knowledge of magnetospheric phenomena.

In the immediate future our plans are to complete a new electron model, AE-7, which will correct some of the quantitative deficiencies shown by the new data available: outer zone OV3-3, OV1-19, AZUR, ATS 6, and OGO 5. In summary, these deficiencies occur in the regions of large gradients, namely, at low altitude, energies above 2 MeV, and in the region around  $L = 3$ , where the large impulsive injection events of magnetic substorms cannot properly be handled, even if one works with yearly averages. It was pointed out earlier that no low-altitude data were analyzed for the production of AE-4. The cut-off values above  $L = 2.4$  are somewhat in error, as can be seen in Figure 8, where a comparison is made between the model and OV3-3 data. Using the data from OV3-3, OV1-19, and AZUR much better cutoffs can be determined. Data from OV1-19, AZUR, ATS 6, and OGO 5 provide the opportunity to extend the models to about 5 MeV. Because the calibration of instruments has improved in recent years, more accurate values are expected.

The following three energetic electron regions are considered:  $R_1$ ,  $L \leq 1.8$ ;  $R_2$ ,  $1.8 < L \leq 5$ ; and  $R_3$ ,  $L > 5$ . It is fairly easy to characterize the improvements expected using these regions. In  $R_1$  there is no longer a Starfish residue and the solar cycle variation has already been modeled, although extension to higher energies should be possible. In  $R_3$  the major improvement will be in the high-energy position and in the cutoff values. The major problem lies in  $R_2$ . Although better cutoff values will be obtained, the long term averages do not show an orderly behavior. This is shown in Figure 9, where the comparison between the time-averaged, 1.7-MeV data and AE-4 are compared. Notice that for  $L < 3.3$ , the values of the two satellites diverge sharply. Yet in this region, during the 10 months spanned by OV3-3 and the 11 months spanned by OV1-19 data, there were two large injection events in each. Consequently, the order of magnitude difference between OV3-3 and OV1-19 at  $L = 3.0$  can only be attributed to the size of the infrequent events observed. Based on this behavior, accurate predictions are impossible. For AE-7 this region will be handled by determining the best time-averaged flux to use in this region for average solar cycle conditions. Then a large injection event model, integrated over time to provide a fluence, will be developed so that the effects of any number of such substorms can be added. The frequency of large events over the solar cycle will be provided.

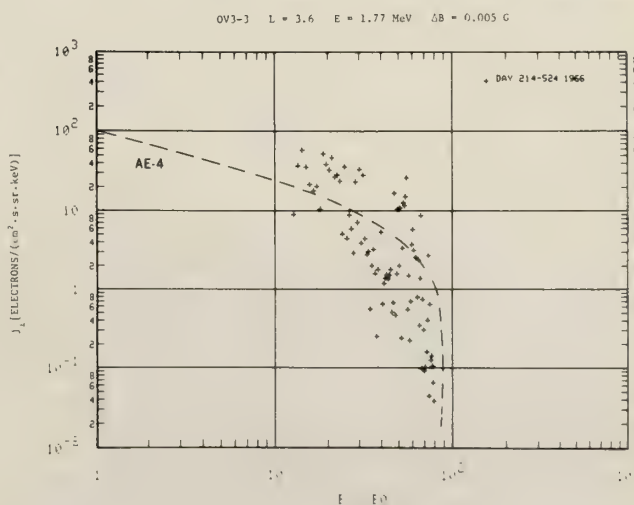


Figure 8. Comparison of low-altitude electron flux cutoff from OV3-3 and model AE-4. The data are daily averages within the indicated B cell; only locally mirroring fluxes are used.



Following the completion of AE-7, a more definitive study of  $R_2$  will be undertaken to determine the non-impulsive or quiet-time injection rate so a better understanding of the solar cycle variation can be obtained. We have examined a number of large substorm injections and find that the fluxes decay exponentially with a time constant,  $\tau$ , that is independent of the event but depends upon  $E$  and  $L$ . An example is shown in Figure 10. Using this, it can be seen that the loss mechanisms are fairly constant in time. This gives a loss rate of  $F/\tau$ , where  $F$  is the instantaneous value of the flux. If the injection rate is denoted by  $Q$ , then

$$\frac{dF}{dt} + \frac{F}{\tau} = Q \quad (4)$$

Although the large events are easy to discern, such a model would indicate that any time the normal decay is not seen, small injections must be occurring. The OV1-19 data of Vampola, during the solar maximum period, offer some insight into the behavior for the  $L$  values inside the synchronous orbit. This data are shown in Figure 11. If one observes the 0.54-MeV electrons at  $L = 3.5$ , 6 impulsive events can be seen. However, there is considerable departure from an exponential decay some 10 or more days after the event. This denotes the occurrence of smaller injections in our view. A number of these can be seen at  $L = 4.5$ . At  $L = 5.5$ , it is nearly impossible to see the large injection events, probably because  $Q$  has a large steady-state value. At  $L = 2.4$ , only three, or possibly four, of the injection events are seen. None of these penetrated past  $L = 2.1$ , although this is not shown in Figure 11. As one looks at 2.6 MeV, one finds that the enhancements disappear at  $L = 2.4$ ,

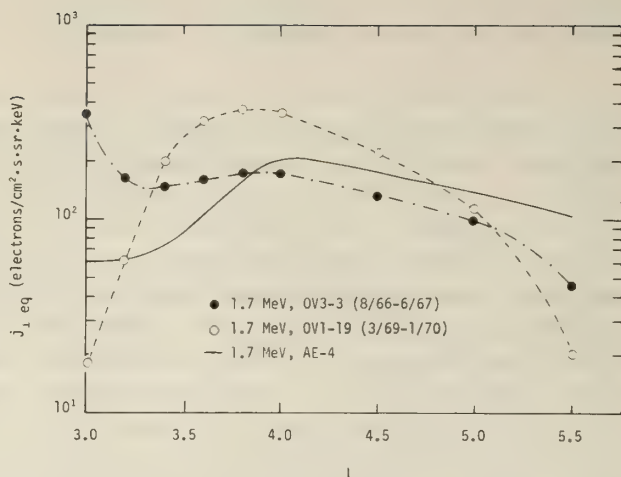


Figure 9. Comparison of long term averages of OV3-3 and OV1-19 data and AE-4 as functions of  $L$ . The fluxes are converted to equatorial mirroring fluxes using the  $B$  distribution of AE-4, using only measurements for  $B < 0.3$  gauss.

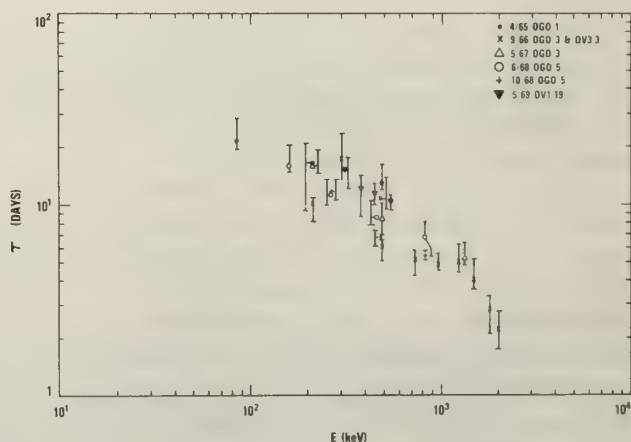


Figure 10. Substorm depletion times at  $L = 2.4$ .

and only three of the six are discernible at  $L = 3.5$ . At  $L = 4.5$  and  $5.5$ , the two large events can barely be distinguished. At  $5.1$  MeV, the two large events are seen at  $L = 3.5$ , but drop to the noise level at  $L = 2.4$  and nearly so at  $L = 4.5$ . It is believed that the extraction of  $Q(\tau)$  from this and other data should lead to some further understanding of the propagation effects of sub-storm injections throughout the trapping regions.

With the confirmation that the OV1-19, AZUR, and OGO 5 data give to AE 6 (Teague *et al.*, 1978) a historical model of the inner zone below  $L = 1.8$  is feasible. Such a model will be constructed using the Starfish decay model and the improved solar cycle variation. This should provide a reasonable description of the whole Starfish decay process.

Besides the real time monitoring available in the geostationary region using the GOES series of satellites, the results recently presented by Baker *et al.* (1978) concerning the change in the pitch angle distribution of electrons between  $30$  and  $300$  keV prior to magnetic sub-storm injections, offer a prediction method for typical times of  $90$  minutes. They observed that the pitch angle distribution changed from one peaked at  $90^\circ$  to one peaked along the magnetic field line in the midnight sector. Their statistical results on the time difference between this peaking parallel to B and sub-storm particle injection time are given in Figure 12. Consequently,  $50$  percent of the time, the warning time prior to particle injection is  $95$  minutes or more.

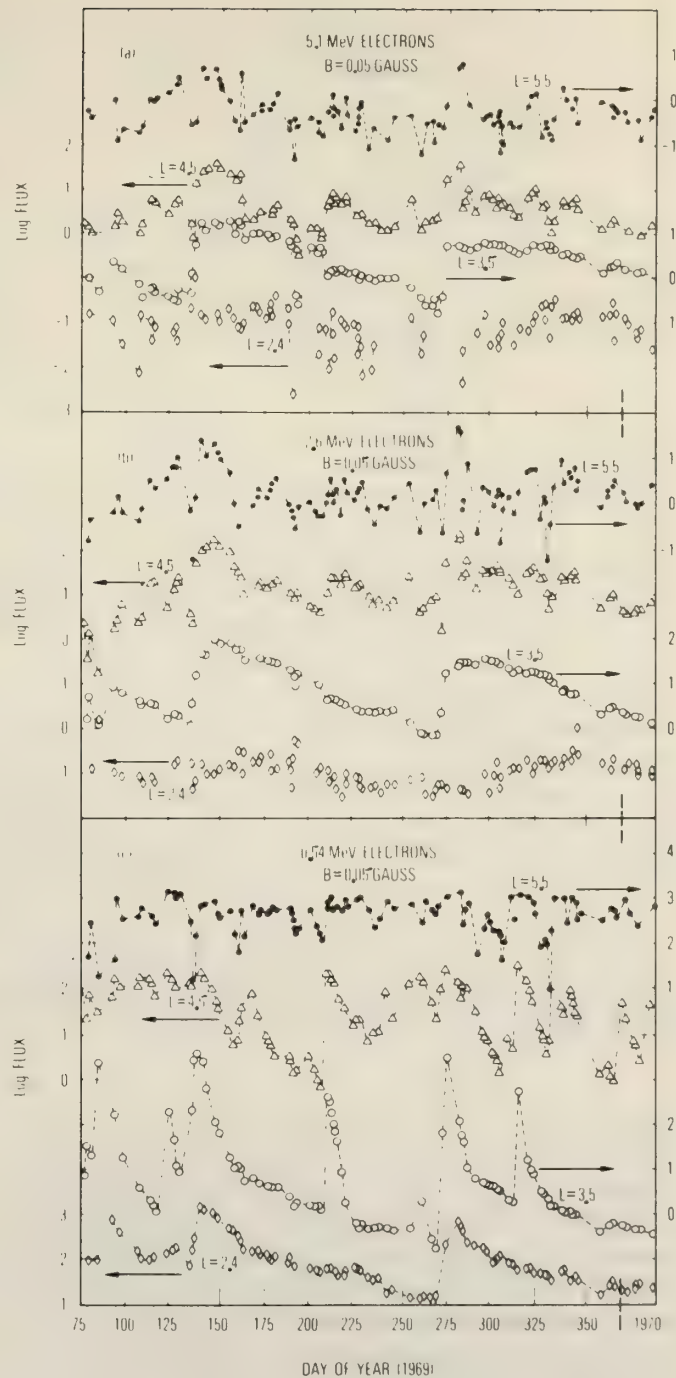


Figure 11. Daily average fluxes from OV1-19 for the period March 1969 through February 1970. Data are all normalized to  $B = .05$  gauss (from Vampola *et al.*, 1977).



The ability to predict injection events offers the possibility of maximizing the data return from satellites below synchronous altitude in order to study the changes that occur in the magnetosphere during such events. Considerable progress toward coordinated data acquisition has been achieved during the International Magnetospheric Study.

We wish to point out another short term prediction possibility that should be useful. The level of electron fluxes at geostationary orbits, particularly those above several MeV, rise following the arrival of fast solar wind streams at the magnetosphere, as demonstrated by *Paulikas and Blake* (1978). Undoubtedly other lower

L values are also effected, but the inner limit has not been determined. With the launch of ISEE 3 on August 12, 1978, and its successful insertion into a halo orbit about the L1 Lagrangian point approximately  $1.5 \times 10^6$  km upstream from the Earth, sensors on board can detect such streams about 30 minutes prior to their arrival at the magnetopause. This would require a real time data display, which was not planned for this mission.

If either or both prediction possibilities were realized, there would be increased emphasis on the modeling of individual events, including the energetic particle population responses.

For the protons, the next level of refinement lies in studying the time variations in quantitative detail. These variations are generally much smaller than those observed in the electron fluxes, but still need to be accounted for in order to improve the accuracy of the models. The time coverage of the data used in constructing AP-8 (shown in Figure 13) leaves something to be desired. Additional proton data that are now available for study are the ISIS 1 data from February 1969 through December 1969, ISIS 2 data from April 1971 through March 1974, and S<sup>3</sup> data from November 1971 through March 1973. The latter data only covers the range from 25 to 875 keV while the ISIS data extends to about 30 MeV. It is known that the lower energy protons exhibit the largest temporal variations, therefore, these should be studied first.

Although the number of experiments flown in the magnetosphere was large, the time-space coverage of this vast volume of space was not very complete. In addition, some instruments in the earlier periods were not very sophisticated and their calibrations left something to be desired. The variety of problems encountered in modeling the trapped radiation environment was discussed recently by *Vette et al.* (1978). Because the trend in magnetospheric studies has been toward the boundary and tail regions in the energetic plasma energy range, the future outlook for good coverage in space, time, and energy

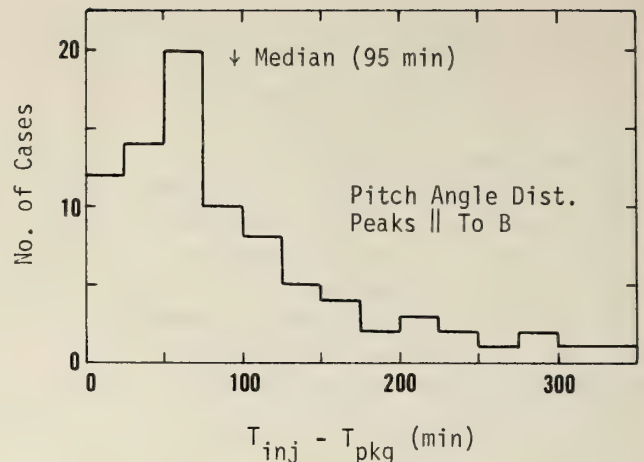


Figure 12. Statistical relationships between the time the electron pitch angle distribution peaks parallel to B,  $T_{pkg}$ , and the time electrons are injected,  $T_{inj}$  (*Baker et al.*, 1978).

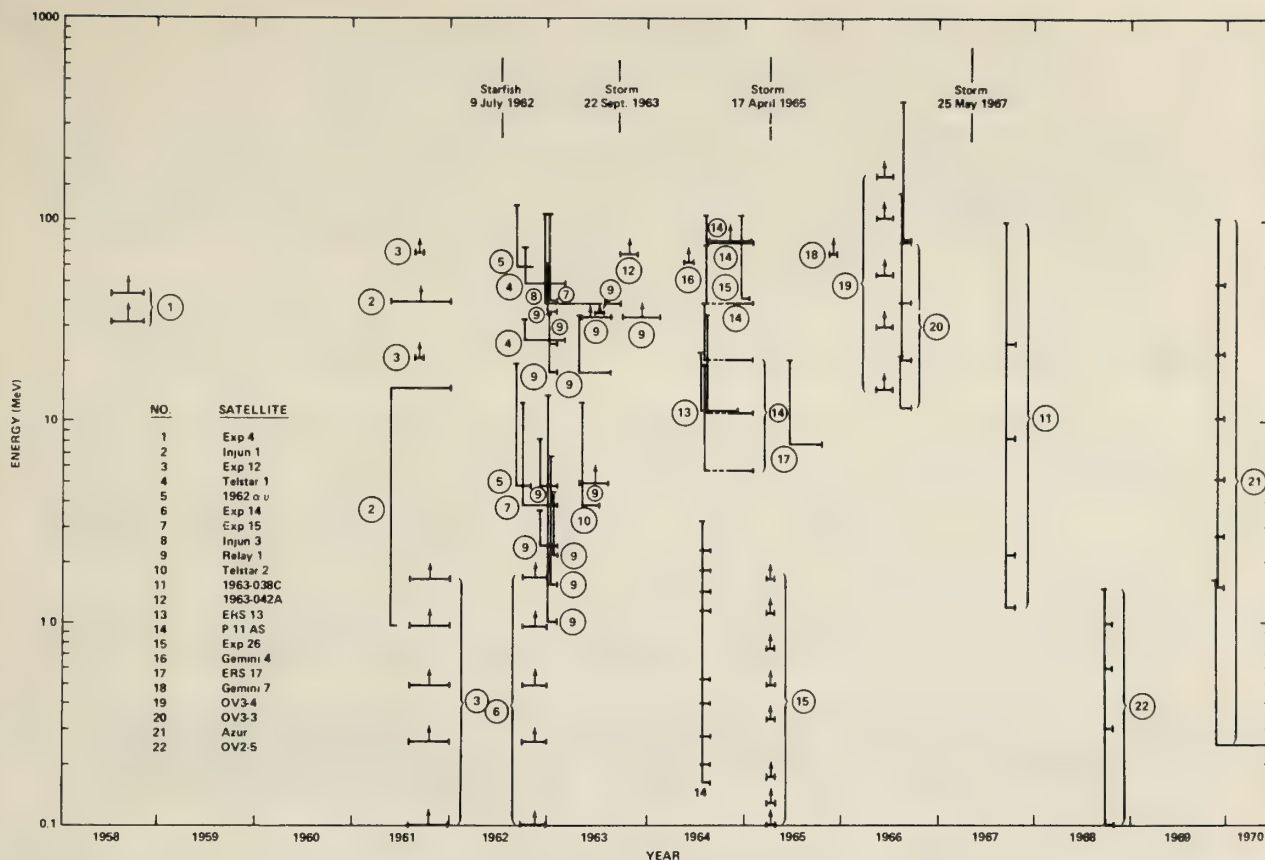


Figure 13. Energy-time coverage for experiments used in constructing the AP-8 model. The nominal energy of each threshold detector is shown with an upward pointing arrow, and the energy bands of interval detectors are represented by the vertical bars. Three major magnetic storms and the Starfish detonation are marked.

is not encouraging. Some data will be available from ATS 6, 1976-059, 1977-007, SMS/GOES and GEOS 2 at synchronous altitudes, and for other regions from ISEE 1 and 2, S3-2, S3-3, and GEOS 1. However, these data do not cover the upper energy ranges with the exception of the ATS 6 data of Paulikas and Blake already discussed. T. A. Fritz (private communication) has noted that there may be some problems with AP-8 in the energy region around 50 keV based on his analysis of S<sup>3</sup> data. The differences may be explained by time variations coupled with the inability of earlier proton experiments to discriminate against alpha particles; however, a detailed analysis must be made before these disagreements can be resolved. Near real time monitoring is available from the SMS/GOES series. Daily summary data are available in hardcopy from GEOS 1 and 2, and in tape and microfilm output from ISEE 1 several months after real time. However, the detailed data from most of the satellites mentioned above, which are necessary for modeling purposes, will not be available for a number of years. Consequently, what was outlined above should provide a reasonable picture of the modeling efforts during the next few years.



## REFERENCES

- Baker, D. N., P. R. Higbie, E. W. Hones, Jr., and R. D. Belian (1978): High-resolution energetic particle measurements at 6.6 R , 3, low-energy electron anisotropies and short term substorm predictions. Los Alamos Scientific Laboratory, University of California Report LA-UR 78-392, Los Alamos, NM 87545. Submitted to J. Geophys. Res.
- Dragt, A. J. (1971): Solar cycle modulation of the radiation belt proton flux. J. Geophys. Res., 37:179.
- King, J. H. (1967): Models of the trapped radiation environment, Vol. IV: low-energy protons. NASA SP-3024, Washington, D.C.
- Lavine, J. P., and J. I. Vette (1969): Models of the trapped radiation environment, Vol. V: inner belt protons. NASA SP-3024, Washington, D.C.
- Lavine, J. P., and J. I. Vette (1970): Models of the trapped radiation environment, Vol. VI: high energy protons. NASA SP-3024, Washington, D.C.
- Paulikas, G. A., and J. B. Blake (1971): The particle environment at the synchronous altitude, in models of the trapped radiation environment, Vol. VII: long term time variations. NASA SP-3024, Washington, D.C., pp 49-67.
- Paulikas, G. A., and J. B. Blake (1978): Energetic electrons at synchronous altitude 1967-1977. Space and Missile Systems Organization Report SAMSO-TR-78-75, July 1978, Los Angeles, CA 90009.
- Roberts, C. S. (1965): On the relationship between the unidirectional and omnidirectional flux of trapped particles on a magnetic line of force. J. Geophys. Res., 70:2517.
- Sawyer, D. M., and J. I. Vette (1976): AP-8 trapped proton environment for solar maximum and solar minimum. National Space Science Data Center Report NSSDC/WDC-A-R&S 76-06, December 1976, Greenbelt, MD 20771.
- Singley, G. W., and J. I. Vette (1972a): A model environment for outer zone electrons. National Space Science Data Center Report NSSDC 72-13, December 1972, Greenbelt, MD 20771.
- Singley, G. W., and J. I. Vette (1972b): The AE-4 model of the outer radiation zone electron environment. National Space Science Data Center Report NSSDC 72-06, August 1972, Greenbelt, MD 20771.
- Stassinopoulos, E. G., and P. Verzariu (1971): General formula for decay lifetimes of Starfish electrons. J. Geophys. Res., 76:110.
- Teague, M. J., K. W. Chan, and J. I. Vette (1976): AE 6: A model environment of trapped electrons for solar maximum. National Space Science Data Center Report NSSDC/WDC-A-R&S 76-04, May 1976, Greenbelt, MD 20771.

- Teague, M. J., N. S. Schofield, K. W. Chan, and J. I. Vette (1978): A study of inner zone electron data and their comparison with trapped radiation models. To be published as National Space Science Data Center Report.
- Teague, M. J., and E. G. Stassinopoulos (1972): A model of the Starfish flux in the inner radiation zone. NASA/Goddard Space Flight Center Report X-601-72-487, December 1972, Greenbelt, MD 20771.
- Teague, M. J., J. L. Stein, and J. I. Vette (1972): The use of the inner zone electron model AE-5 and associated computer programs. National Space Science Data Center Report NSSDC 72-11, November 1972, Greenbelt, MD 20771.
- Teague, M. J., and J. I. Vette (1972): The inner zone electron model AE-5. National Space Science Data Center Report NSSDC 72-10, November 1972, Greenbelt, MD 20771.
- Teague, M. J., and J. I. Vette (1974): A model of the trapped electron population for solar minimum. National Space Science Data Center Report NSSDC 74-03, April 1974, Greenbelt, MD 20771.
- Vampola, A. L., J. B. Blake, and G. A. Paulikas (1971): A new study of the magnetospheric electron environment. J. Spacecraft and Rockets, 14: 690.
- Vette, J. I. (1966): Models of the trapped radiation environment, Vol. I: inner zone protons and electrons. NASA SP-3024, Washington, D.C.
- Vette, J. I., A. B. Lucero, and J. A. Wright (1966): Models of the trapped radiation environment, Vol. II: inner and outer zone electrons. NASA SP-3024, Washington, D.C.
- Vette, J. I. and A. B. Lucero (1967): Models of the trapped radiation environment, Vol. III: electrons at synchronous altitudes. NASA SP-3024, Washington, D.C.
- Vette, J. I., K. W. Chan, and M. J. Teague (1978): Problems in modeling the Earth's trapped radiation environment. Air Force Geophysics Laboratory Report AFGL-TR-78-0130, Hanscom AFB, MA 01731.



# LOW ENERGY MAGNETOSPHERIC PLASMA INTERACTIONS WITH SPACE SYSTEMS - THE ROLE OF PREDICTIONS

H. B. Garrett  
Air Force Geophysics Laboratory  
Hanscom AFB, Bedford, MA 01731

The present status of low energy magnetospheric plasma interactions with space systems is reviewed. The role of predictions in meeting user needs in assessing the impact of such interactions is described. In light of the perceived needs of the user community and of the current status of modeling and prediction efforts, it is suggested that for most user needs more detailed statistical models of the low energy environment are required. In order to meet current prediction requirements, real-time in situ measurements are proposed as a near-term solution.

## INTRODUCTION

As we pass from the age of exploration of the near-earth space environment to the exploitation phase, an increasing need has arisen for an understanding of the environmental interactions between space systems, space systems operations, and the low energy (0-100 keV) plasma environment. The problem of concern to the space physics community remains, however, the definition of this environment, and how it is perturbed by natural and, increasingly, man-made variations. It has become clear, however, that a careful choice must be made in deciding what parameters - both descriptive and predictive - are necessary to our understanding of the near-earth regime. An important element in this choice that has been largely ignored is the issue of user needs. Much of this review will be based on the perceived needs of the existing and potential user community in this era of space exploitation of the near-earth magnetospheric environment.

In this review the present status of low-energy plasma modeling and, where necessary, the related issue of the geomagnetic and magnetospheric electric field modeling will be covered. In light of the growing role of satellite-environment interactions, this area will also form a major part of the review. As an adjunct, current attempts at predicting geophysical and,

hence, magnetospheric variations by employing various geomagnetic indicators are discussed. An example will then be presented in which all three areas are linked to form a reasonably coherent picture of one phenomena - spacecraft charging in the geosynchronous environment.

The review will conclude with a discussion of important areas for research. In particular, models of a variety of geophysical processes and physical interaction mechanisms are needed in the immediate future. Not only do measurements of the low energy plasma environment need to be expanded, but currently existing data sets must be better organized and exploited. Finally, the necessity for real-time, in-situ measurements to establish the current state of the magnetosphere will be put forward as a near-term alternative to detailed magnetospheric modeling.

### THE NEED FOR LOW ENERGY PLASMA/SPACE SYSTEMS INTERACTION PREDICTION

Unlike the situation for high energy particle modeling where radiation damage is a known threat or for the neutral particle population where atmospheric drag is a well-recognized problem, the effects of the low energy near-earth plasma environment on space systems are at best subtle (see also review by Johnson *et al.*, 1979). It has only been with the advent of the new generation of sophisticated communication and scientific satellites that previously minor systems interactions such as spacecraft charging and surface contamination have become of growing concern to the spacecraft designer and user. In this section the principal interactions that have come to light will be outlined and the associated problem areas from the designer/user standpoint defined.

The principal effect of the low energy particle environment on spacecraft systems is currently believed to be spacecraft charging. Although recognized early as a source of error in low energy particle measurements in the plasmasphere (Whipple, 1965; Whipple and Parker, 1969a, b) and as a potential source of drag (see review by Brundin, 1963), it was not until the observations of significant potentials on the order of -10 KV on the geosynchronous ATS-5 satellite during eclipse (DeForest, 1972) that an interest was really taken in the phenomenon. An example of this phenomenon is given in Figure 1 for the ATS-6 satellite. Likewise, it was not until various satellites (Shaw *et al.*, 1976; McPherson and Schober, 1976) were shown to have suffered what were assumed to be logic upsets resulting from arcing between differentially charged surfaces (Figure 2) and a relationship between arcing and  $a_p$  (Figure 3) found that the user community began to become concerned and sufficient interest generated to support a joint AF/NASA spacecraft charging program. This joint effort recently resulted in the successful launch of the P78-2 (SCATHA) satellite - the first satellite designed to study



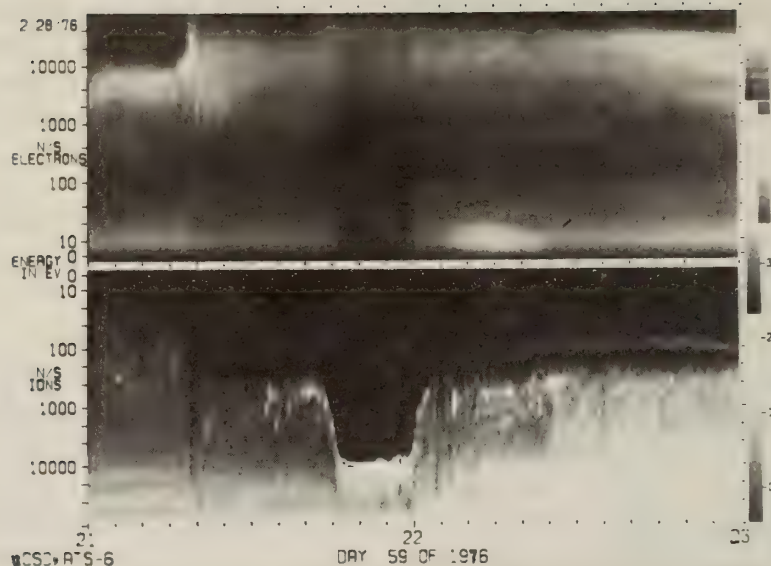


FIGURE 1. Spectrogram of Day 59, 1976 from ATS-6 (see DeForest and McIlwain, 1971, for explanation of scales). The dropout in the ions between 2145-2200 UT is simultaneous with satellite entry into eclipse and reflects in eV the negative potential  $V$  in volts on the satellite as it became charged due to the loss of photoelectrons.

specifically environmental interactions. The buildup of static charge on satellites (particularly at geosynchronous orbit) by the low energy plasma not only affects particle measurements, but is also believed to generate a wake (Parker, 1977 and references therein) which could be a potential source of plasma waves and structural deformation of structurally weak surfaces such as the solar sail (Douglas *et al.*, 1977). Even the passage of a structure in and out of eclipse can pose a threat due to potential gradients and current flows (Gauntt, 1979).

The fact that spacecraft surfaces can become charged also contributes significantly to optical surface contamination. As discussed in Cauffman (1973), this is believed to be tied to the ionization of outgassed contaminants from the spacecraft which then are reattracted to charged surfaces. Likewise, the fact that a spacecraft is immersed in ionized plasma can lead to problems such as multipacting (Freeman and Reiff, 1979) or power losses

## LOCAL TIME DEPENDENCE OF ANOMALIES

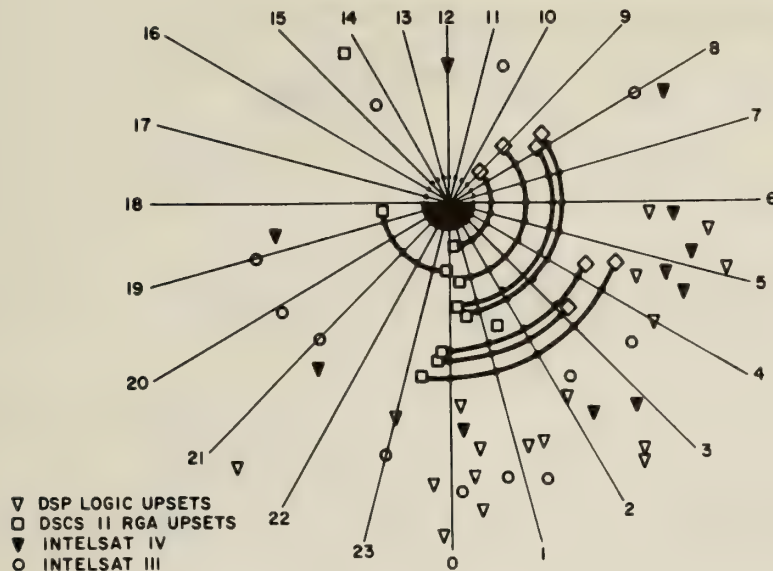


FIGURE 2. Local time dependence of circuit upsets for several DOD and commercial satellites (McPherson and Schober, 1976). Radial position has no significance.

in high voltage solar power panels (McCoy et al., 1979).

As structures grow significantly in size, the problems just discussed grow in effect and in potential for damage. In Figure 4 we have plotted the predicted growth rate of large structures. Such effects as VXB electric fields generated in spacecraft which are of nuisance value now could easily become significant as structural dimensions grow as projected in this figure. More ominous, however, is the potential for such large structures to perturb the environment. At geosynchronous orbit, as an example, Garrett and DeForest (1979a) estimate photoelectron emission rates on the order of  $.4 \text{ nA/cm}^2$  (this is believed to be the flux that actually escapes, not necessarily the total emission rate at the satellite surface). Such fluxes of low energy ( $\sim 10 \text{ eV}$ ) particles are relatively unimportant in the plasmasphere but could



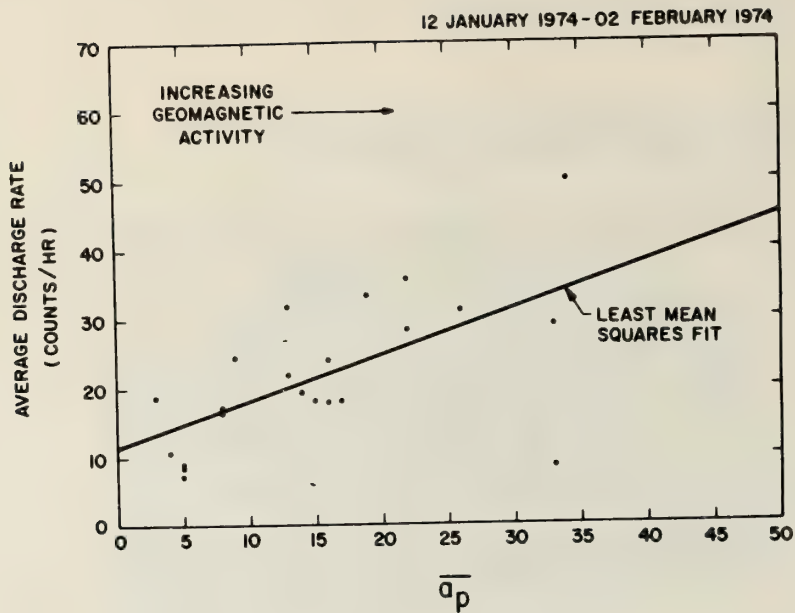


FIGURE 3. Daily average discharge rate as a function of daily average  $a_p$  (from Shaw et al., 1976).

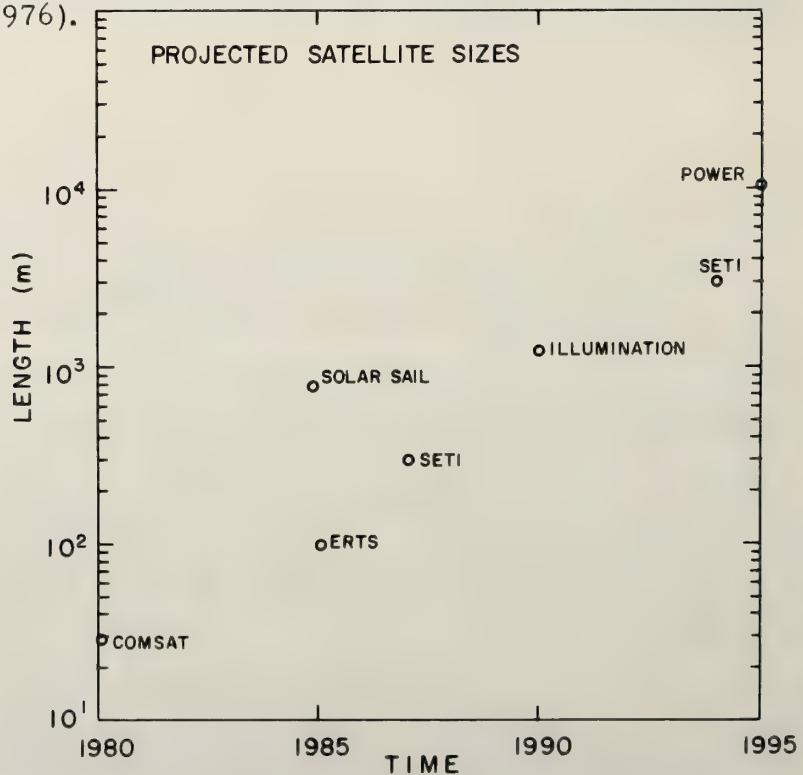


FIGURE 4. Projected maximum dimension of various space systems as a function of time (adapted from Hagler et al., 1977).

be potentially damaging in the plasma sheet where they rival the ambient flux and represent a much colder component (Garrett and DeForest, 1979b).

Contamination of the near-earth environment by photoelectrons is only one of many potentially damaging space system effects on the environment. Contaminant clouds, already discussed in conjunction with degradation of optical and thermal surfaces, also threaten the ambient low-energy plasma environment. Sputtering due to high energy particle impacts, multipacting, chemical exhausts, and, particularly, ion engine exhausts (Luhmann, *et al.*, 1978; Chiu, *et al.*, 1979) all threaten the status quo. Luhmann *et al.* (1978) have shown (Figure 5) how the deposition of large amounts of argon in the plasmasphere due to solar power satellite operations will result in the conversion of the region from a hydrogen dominated to oxygen dominated region. Finally, large structures will absorb correspondingly large amounts of the ambient flux. Although this will probably be more important for the high energy particle environment, measurable perturbations of the low energy plasma are a possibility if extensive operations at geosynchronous orbit are implemented as planned.

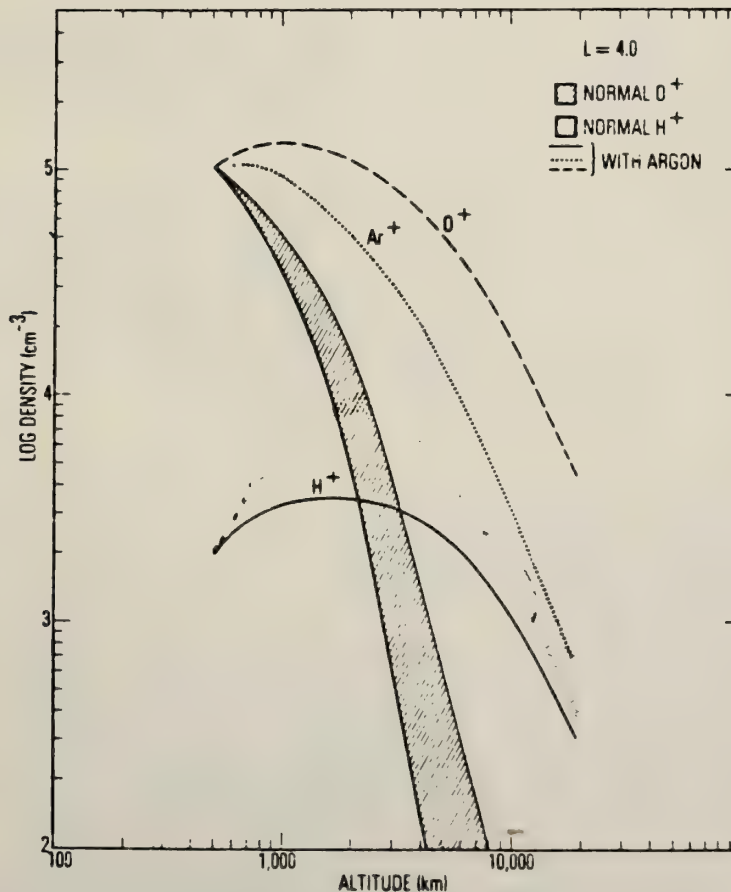


FIGURE 5. Projected variations of  $\text{H}^+$ ,  $\text{O}^+$ , and  $\text{Ar}^+$  before and after the passage of a large solar power satellite to geosynchronous orbit (Luhmann *et al.*, 1978).



Although several papers on the subject of spacecraft interactions have been referenced, the fact remains that at present relatively little is actually known about most of the interaction processes mentioned. The problem is that there do not as yet exist sufficiently detailed models of many of the dynamic processes associated with the movement of magnetospheric plasma. Similarly, the level of precision necessary to define many of the interactions is severely limited and represents a pressing problem. These two problems are not and cannot be divorced from each other as we must define the parameters necessary to accurately represent the environment and at the same time those parameters must be keyed to the interaction models to be of any use. As will become apparent, a real dichotomy exists between the parameters currently used to define the environment and those required to predict plasma interactions.

## PRESENT STATUS

In order to demonstrate the difficulties attendant with studying the problems of low energy particle interactions with space systems, the current status of these efforts will be reviewed. First, a brief review of the current status of low energy magnetospheric plasma models will be given. This will be followed by a discussion of attempts at predicting and modeling the interaction phenomena associated with the low energy environment. As other committees are considering in detail the prediction of magnetic and electric field variations and of geomagnetic indices, only a cursory description will be given of these parameters. Finally, an example will be given of one attempt to link these three areas into a coherent predictive model of one interaction phenomena - spacecraft charging during eclipse passage.

### Low Energy Plasma Models

In Garrett (1979) the current status of the modeling of 0-100 keV mid-magnetospheric particle environment is reviewed in great detail. In this section some of the main results of that study will be presented, the reader being referred to that review for a more detailed description. As in Garrett (1979), comments will be confined to the ionospheric and auroral domains as described by Vasyliunas (1972). That is, only the low energy (0-100 keV) charged particle population in the plasmasphere and the near-earth plasma sheet regions will be considered. Also, variations in ionic composition will be ignored (see Young, 1979, for an excellent treatment of ionic composition variations).

Four types of quantitative models will be described dependent upon the ratio of theoretical to empirical input in the model. The most elementary models to be discussed consist of statistical compendiums of various parameters as functions of space, time, and geomagnetic activity. These sta-

tistical models require little theoretical input, relying primarily on actual measurements. Consideration of basic physical principles makes possible the derivation of analytic expressions capable of simulating changes in the environment - the second type of model. Third are models which employ theory to predict trajectories of particles in static electric and magnetic fields. Finally, the most complete model from a theoretical standpoint is a full, 3-dimensional, time-dependent model capable of taking into account time-varying injection events. The following discussion will center on these four categories of models.

Statistical models, as defined here, are compendiums or histograms of various plasma parameters based on actual data. The basic examples of this type of model are the composite or average distribution functions generated by Chan *et al.* (1977) for various magnetospheric and solar wind regions (Figure 6). Although such descriptions are particularly useful in predicting long term dosages, a prohibitive number of distribution functions are needed as a function of time, spatial coordinates, and geomagnetic activity to adequately describe the near-earth magnetosphere. Instead, a description in terms of the first four plasma moments (density, number flux, pressure or energy density, and energy flux) from which a number of para-

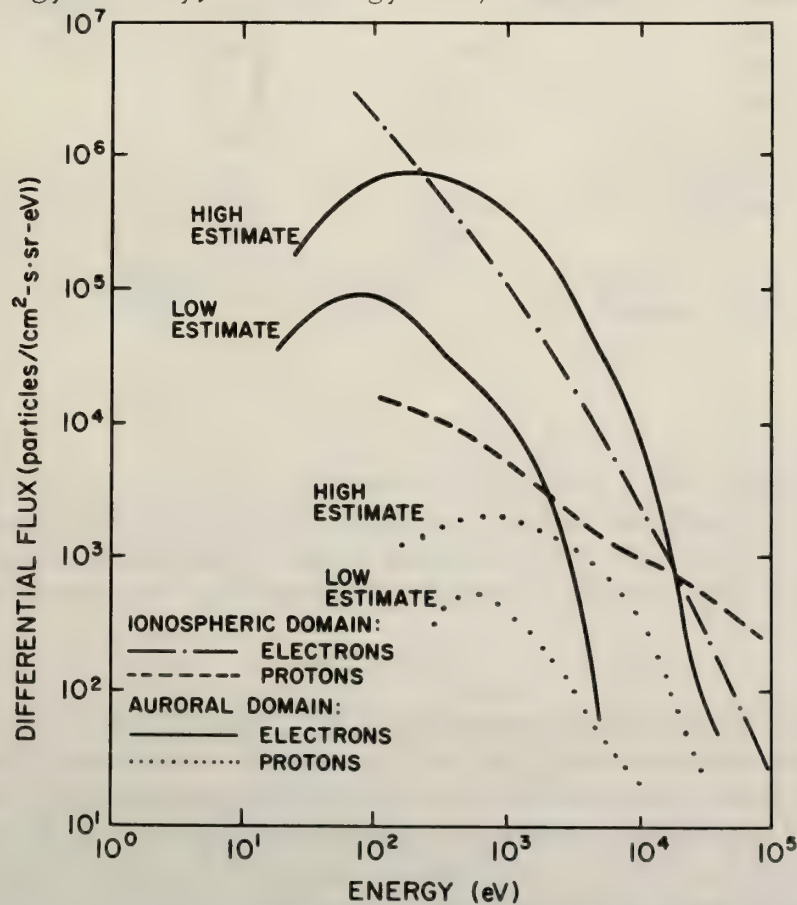


FIGURE 6. Estimated differential particle fluxes for the auroral and ionospheric domains (from Chan *et al.*, 1977).



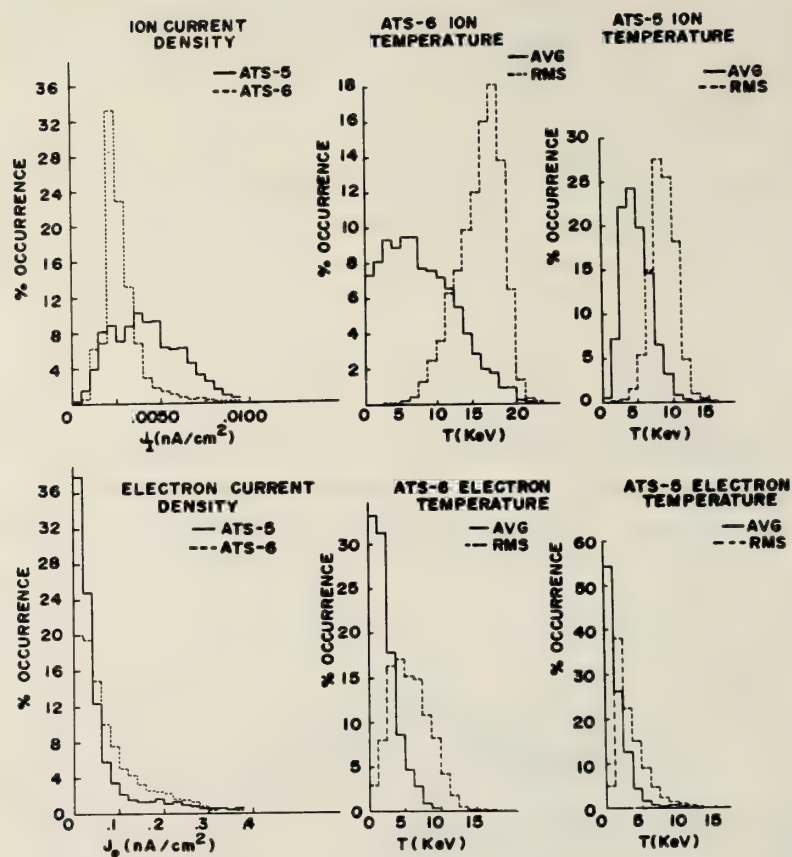


FIGURE 7. Histograms of the occurrence frequencies of the electron and ion temperatures and current at geosynchronous orbit as measured by ATS-5 and ATS-6.  $T(\text{AVG})$  is  $2/3$ 's the ratio of energy density to number density  $T(\text{RMS})$  is one half the ratio of particle energy flux to number flux (Garrett et al., 1978b).

parameters such as current and temperature can be derived, has emerged as a compact means of describing the environment. Vasyliunas (1968), DeForest and McIlwain (1971), Su and Konradi (1977), and Garrett et al. (1978b) have all carried out statistical studies of these parameters. In Figure 7, results from the analysis of ATS-5 and ATS-6 temperatures and currents are plotted as an example ( $T(\text{AVG})$  is the temperature obtained from dividing the energy density by the number density,  $T(\text{RMS})$  is the temperature obtained from dividing the energy flux by the number flux). Models have also been

developed for other spatial positions by plotting the results from eccentric, inclined satellites in the manner of the intensity plots of high energy particles. A particularly good example is given in Figure 8 (Frank, 1967) for the low energy ions ( $200 \text{ eV} \leq E \leq 50 \text{ keV}$ ) in the generalized  $R - \lambda_m$  coordinate system. Similar studies (Carpenter, 1966; Chappell *et al.*, 1970; Lennartsson and Reasoner, 1978; and reviews by Chappell, 1972, and Carpenter and Park, 1973) have also been carried out in the plasmasphere.

The major difficulty with the preceding statistical models from a predictive standpoint is their inability to include the effects of complex time variations while maintaining the interrelationships between the various parameters. One solution to this problem is simply to provide several detailed but representative examples (see DeForest and Wilson, 1976) of actual spectra. Unfortunately, from a practical standpoint such information lacks compactness and does not easily provide many of the parameters required by potential users. A solution to this problem of tradeoff between accuracy and massive amounts of data has been the introduction of analytic equations capable of modeling specific parameter variations. An example is the well known formula for the midnight-dawn plasmopause of Carpenter and Park

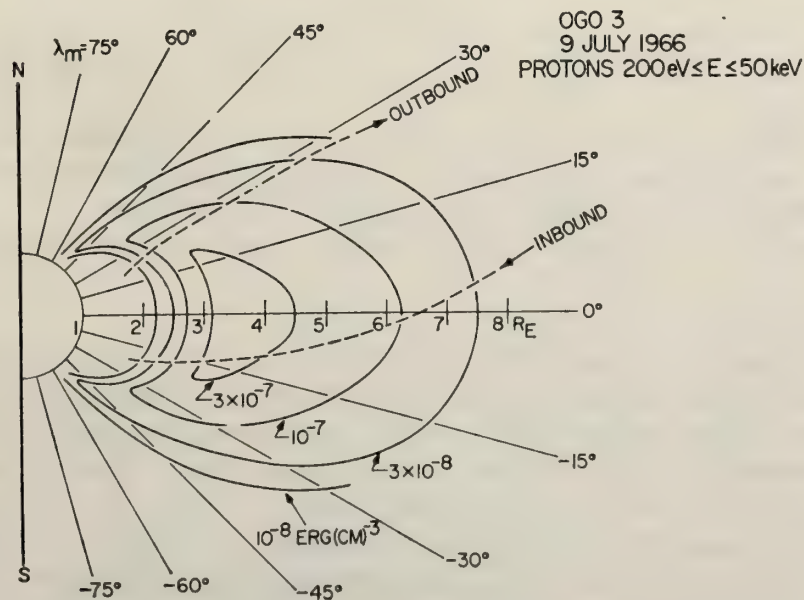


FIGURE 8. Contours of constant proton ( $200 \text{ eV} \leq E \leq 50 \text{ keV}$ ) energy density as measured by OGO 3 on 9 July 1966 in a  $R - \lambda_m$  coordinate system (Frank, 1967).



(1973) (see also Mauk and McIlwain, 1974 ; Freeman, 1974 ; and, particularly the review of injection boundary formulas by Kivelson et al., 1978):

$$L_{pp} = 5.7 - 0.47 K_p \quad (1)$$

where  $L_{pp}$  is the plasmopause boundary (in earth radii) and  $K_p$  is the maximum value of  $K_p$  in the preceding twelve hours.

Fairly detailed analytic formulas have been or are being developed for a variety of plasma parameters as functions of time, spatial position, and geomagnetic activity. Su and Konradi (1977), Garrett (1977), and Garrett and DeForest (1979b) have all developed analytic expressions capable of defining the geosynchronous environment. Garrett and DeForest (1979b) in particular have developed a compact representation in which the first four moments of the distribution function at geosynchronous orbit are expressed in terms of equations linear in  $A_p$  (daily average of  $a_p$ ) and varying diurnally and semi-diurnally in local time, LT:

$$M_i (A_p, LT) = (a_0 + a_1 A_p) (b_0 + b_1 \cos (\frac{2\pi}{24} (LT + t_1)) + b_2 \cos (\frac{4\pi}{24} (LT + t_2))) \quad (2)$$

where:  $M_i$  = moment  $i$

$a_0, a_1, b_0, b_1, b_2, t_1, t_2$  = fitted parameters.

Typical results are given in Figure 9. A major advantage of this model is that a 2-Maxwellian (i.e., the sum of 2 distinct plasma components) distribution function can be derived directly from the data. Konradi (private communication) is extending this model to lower altitudes and different latitudes.

The previous two model types provide most of the data needed in predicting environmental effects on space systems. In order, however, to fully understand the effects of space systems on the environment, a rudimentary knowledge of how man-made contaminants and perturbations of the environment propagate in the magnetosphere is required. Thus, static models of the magnetospheric fields play an important role in this review. Unlike the preceding models, considerable work has gone into studying such models so that only a few representative examples will be given here (please see review by Garrett, 1979).

The most typical example of static field models is that developed by McIlwain (McIlwain, 1972) in conjunction with the ATS-5 geosynchronous data. By careful analysis of the data, McIlwain constructed electric and magnetic field models capable of reproducing the observed spectra. Some

# ELECTRON TIME VARIATIONS

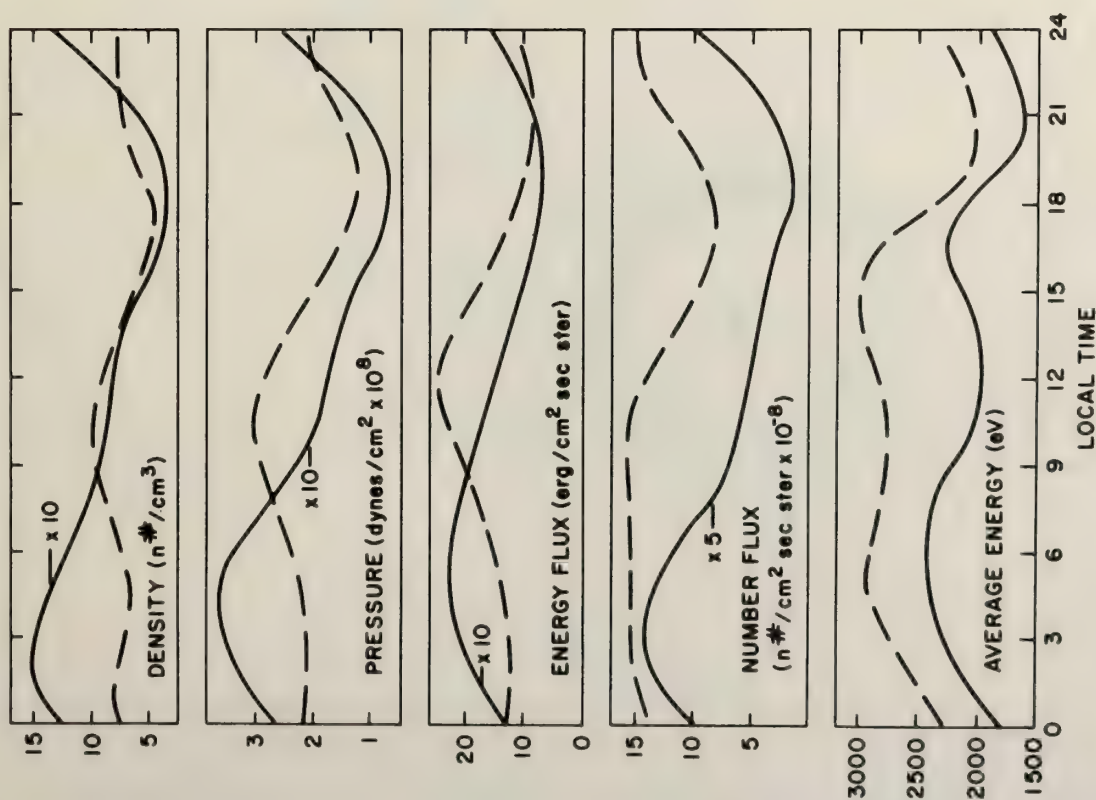


FIGURE 9a. Electron parameter time variations as a function of local time (Garrett et al., 1979a). The solid line is for  $A_p = 15$  and the dashed line for  $A_p = 207$ .

# ION TIME VARIATIONS

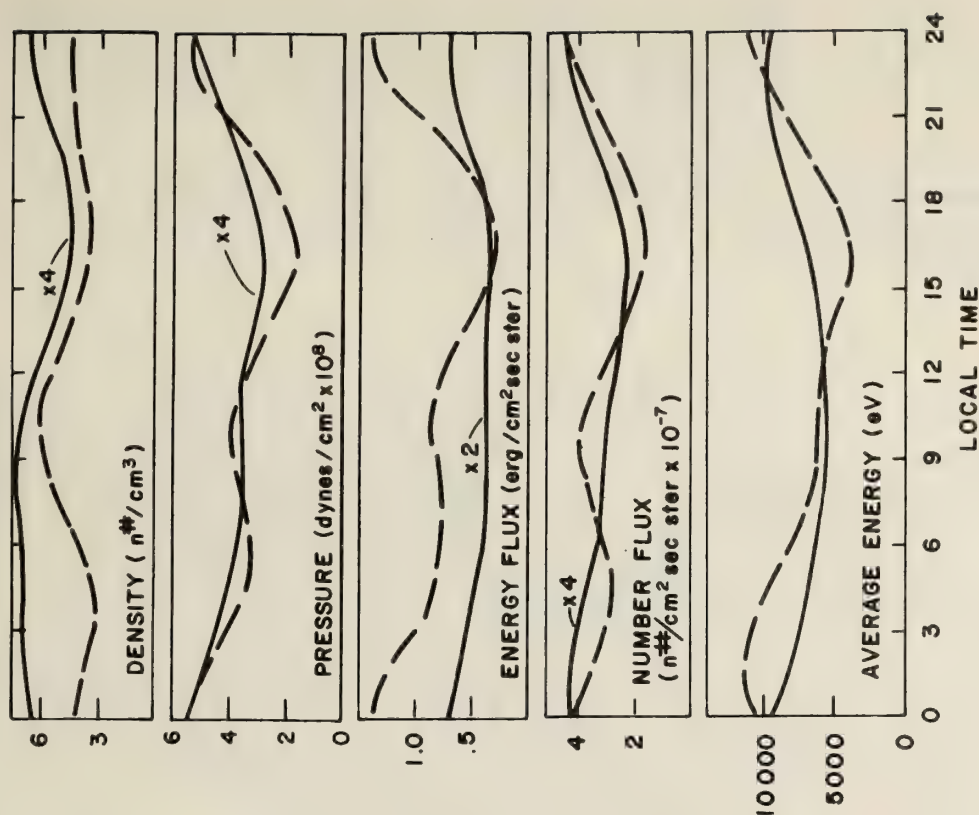


FIGURE 9b. Ion parameter time variations as a function of local time (Garrett et al., 1979a). The solid line is for  $A_p = 15$  and the dashed line for  $A_p = 207$ .



results of his analysis for typical electron and ion trajectories in his model fields are given in Figures 10a and 10b. Many variations exist on McIlwain's model ranging from the early models of Kavanagh *et al.* (1961), Roederer and Hones (1970), and Wolf (1970) down to the more recent work of Konradi *et al.* (1975) and Walker and Kivelson (1975). The major difference between most of these models is the exact treatment of the plasma injection event (see review by Kivelson *et al.*, 1978). The resolution of this problem is dependent on what role the in situ and ionospheric plasma sources play and is still controversial.

The static models of the plasmasphere divide into two principal categories - ionospheric diffusion and EXB drift. By way of explanation, Chappell *et al.* (1970) and, to a lesser extent, Wolf (1970) and Chen (1970) model the drifts of particles in the plasmasphere in a similar fashion to McIlwain - that is, by EXB drift. Others such as Angerami and Thomas (1964), Schunk and Walker (1969), and Mayr *et al.* (1973) model the diffusion of ionospheric particles into the exosphere and plasmasphere. Recently, more detailed models by Lemaire and Scherer (1974) and Chiu *et al.* (1978) have produced comprehensive models in this latter category by using a collisionless kinetic theory to compute the diffusion along the field line of various constituents.

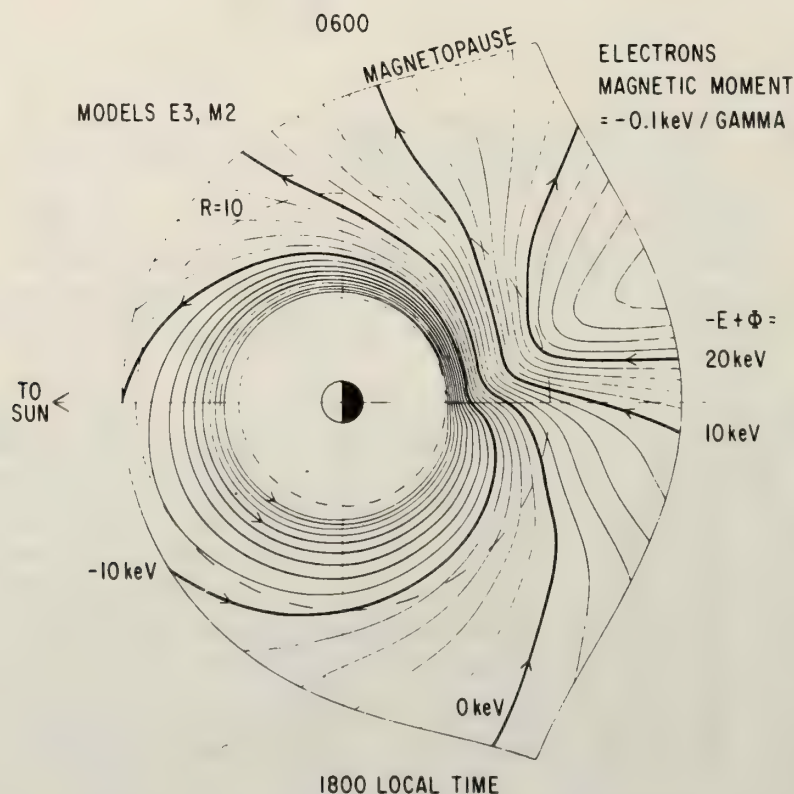


FIGURE 10a. Trajectories for electrons having  $\mu = 0.1 \text{ keV}/\gamma$  in the McIlwain model (McIlwain, 1972).

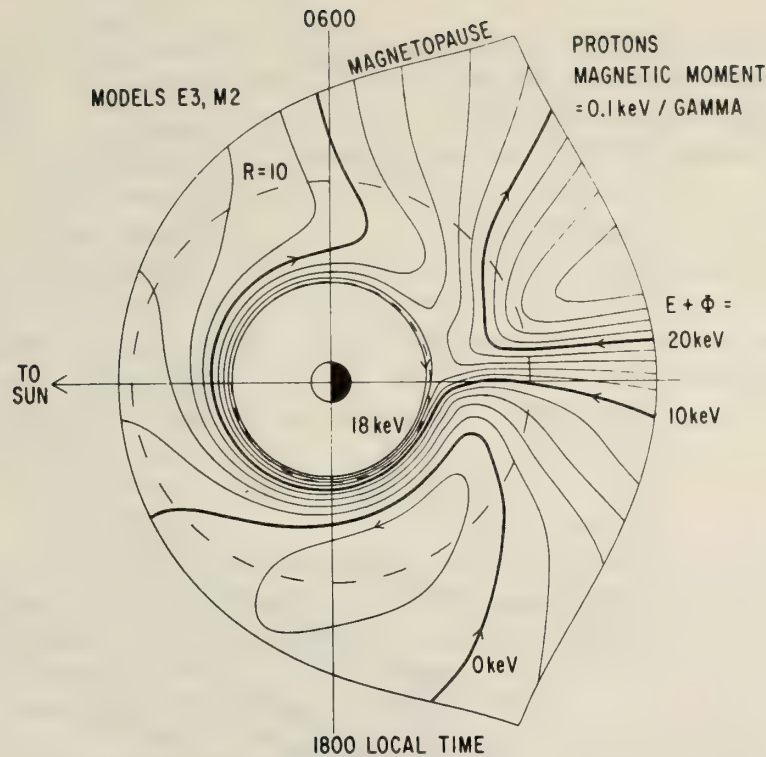


FIGURE 10b. Trajectories for protons having  $\mu = 0.1 \text{ keV} / \gamma$  in the McIlwain model (McIlwain, 1972).

Although most of the effort in magnetospheric modeling has gone into static models, an increasing amount of detailed work on the time-dependent behavior of the low energy plasma environment has appeared. Specifically, Roederer and Hones (1974) have successfully reproduced many of the features of the ATS-5 data by assuming a static field model upon which they have superimposed a time-varying electric field component. In a similar vein, Smith *et al.* (1978a, 1978b, 1979) have used a time-varying convection electric field potential of the Volland-Stern type (Ejiri *et al.*, 1977):

$$\phi = AR^2 \sin \phi \quad (3)$$

where:

$\phi$  = local time dependence (0 at midnight)

R = radial distance from earth

A is taken to be of the form (Grebowsky and Chen, 1975):

$$A = \frac{0.045}{(1 - 0.159 K_p + 0.009 K_p^2)^3}$$



Assuming a dipolar magnetic field, Smith et al. (1978a, 1978b, 1979) then produced a movie of the plasma flow entitled "Convection of Magnetospheric Particles in a Time-Varying Electric Field."

The final models to be discussed are the detailed 3-dimensional, time-dependent models of the Rice University group under R. A. Wolf. Chen (1970), Wolf (1970), Chen and Wolf (1972), Jaggi and Wolf (1973), Wolf (1974), and more recently Southwood (1977) and Harel et al. (1978) are papers in this series. Unique among magnetospheric models, they calculate the particle drifts and the resulting effects on the electric and magnetic fields in a self-consistent fashion. Some of their results for the magnetosphere (Jaggi and Wolf, 1973) and the plasmasphere (Chen and Wolf, 1972) are presented in Figures 11a, 11b, and 11c. Note in particular how the plasma "tails" off from the plasmasphere. As this phenomena represents a loss mechanism for the plasmasphere, it may be important for calculations of contaminant lifetimes in this region.

### Interaction Modeling

As currently practiced, modeling or prediction of interactions with the low-energy plasma environment is synonymous with spacecraft charging. As a result most of the discussion of this section will be concerned with current

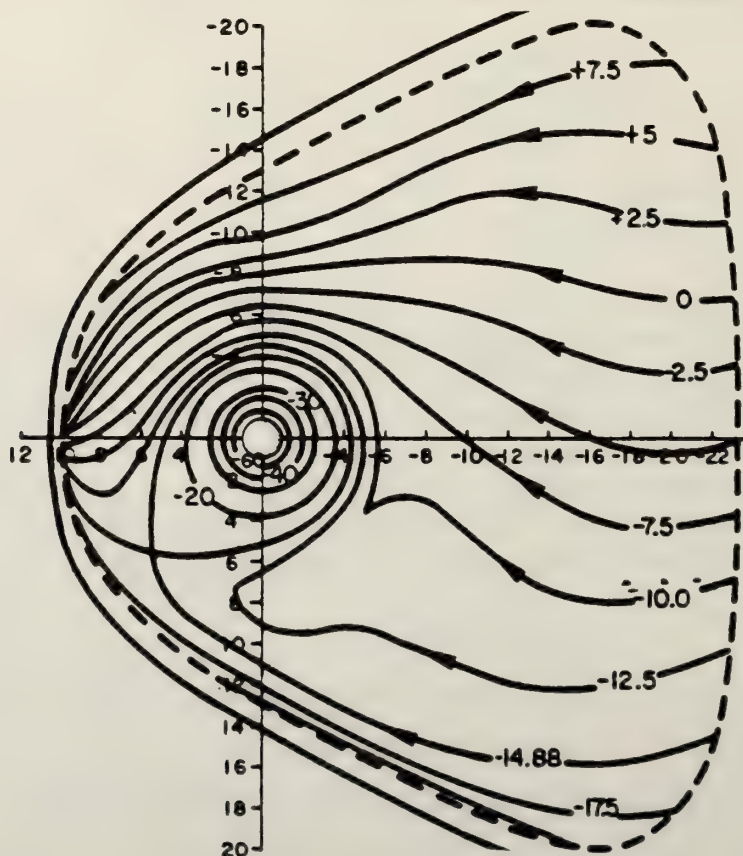


FIGURE 11a. The electrostatic potential 6 hours after a sheet of ions started moving toward the earth from the magnetospheric tail (Jaggi and Wolf, 1973).

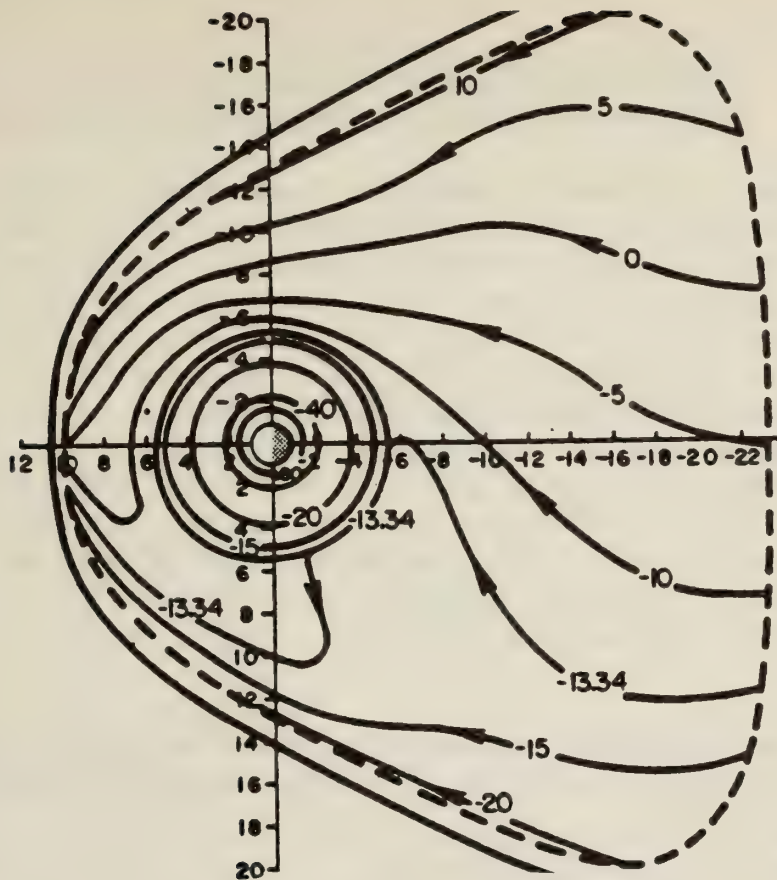


FIGURE 11b. The electrostatic potential 15 hours after the sheet started moving towards the earth from the magnetospheric tail (Jaggi and Wolf, 1973).

efforts at modeling the charging phenomena and related problems such as contaminant effects. In the long term, however, the prediction of environmental effects may likely be more critical. Thus, the final part of this section will briefly review the few attempts in this area to date.

Spacecraft charging is, simply, the attempt on the part of the currents to the spacecraft surface to come to equilibrium:

$$J_E - (J_I + J_{SE} + J_{BSE} + J_{PH} + J_X) = 0 \quad (4)$$

where:

- $J_E$  = Incident electron current
- $J_I$  = Incident ion current
- $J_{SI}$  = Secondary electron current due to ions
- $J_{SE}$  = Secondary electron current due to electrons
- $J_{BSE}$  = Backscattered electron current
- $J_{PH}$  = Photoelectron current (zero when a surface is shadowed.)



$J_X$  = Thruster currents and other miscellaneous currents  
(normally assumed to be zero in simple calculations).

All of the currents are complex functions of spacecraft potential, the satellite plasma sheath (the region over which the satellite perturbs the ambient medium), the satellite materials, and the physical structure of the satellite. Several models (see Whipple, 1965; Rothwell *et al.*, 1977; Laframboise and Prokopenko, 1977; Parker, 1977; and references therein) carry out fairly detailed calculations of the particle trajectories in the vicinity of a satellite allowing simulations of the sheath and differential potentials. A variety of simple single point (or "thick sheath") models also exist which allow approximate calculation of the equilibrium potential on spacecraft (see Garrett and Rubin, 1978, and references therein). These models have been extended by the inclusion of equivalent resistances and capacitances (Inouye, 1976; Massaro *et al.*, 1977; Gauntt, 1979) that actually allow the simulation of the time response of satellites to changes in the ambient particle and photon environment.

In all of the above, however, the detailed shape of the sheath around the spacecraft is not calculable except for some rather simple geometries.

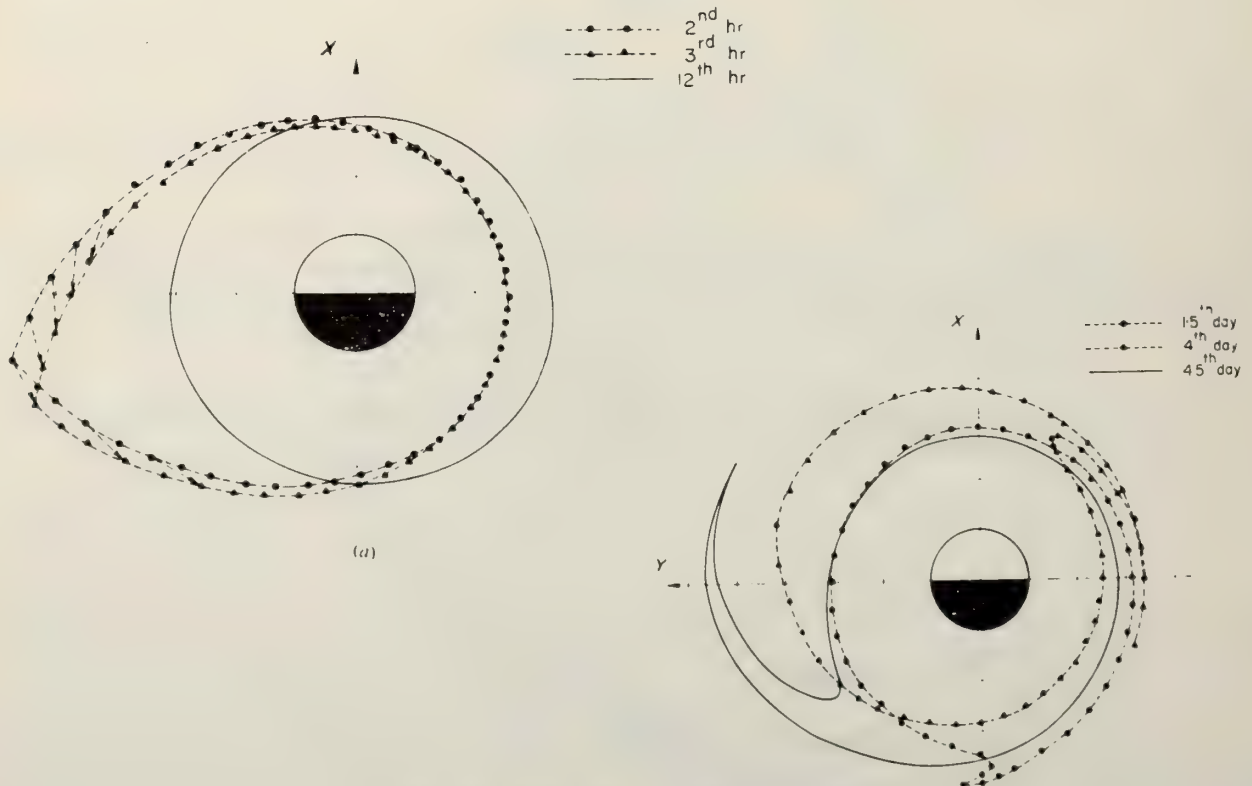


FIGURE 11c. The distortion of the plasmopause at various times following a factor of two decrease in the convection field at time  $t = 0$  (Chen and Wolf, 1972).

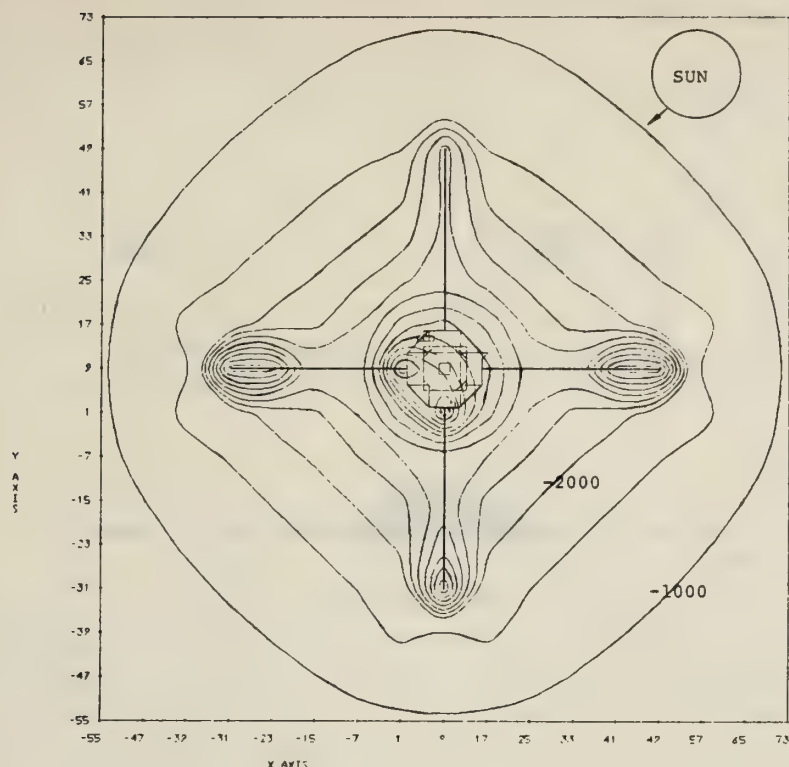


FIGURE 12. Potential contours for sunlit case in a horizontal plane through the SCATHA satellite center. Contours are at 500 volt steps (the minimum is -7500 volts). From Schnuelle et al. (1979).

Recently, the S<sup>3</sup> Corporation of California developed a detailed computer code capable of simulating a complex satellite including booms and shadowing (Katz et al., 1979). The results from one of these calculations is presented in Figure 12 where the potentials around a simulation of the SCATHA satellite are plotted. The code is currently undergoing intensive comparison with data from laboratory experiments and from the SCATHA satellite. It is planned that in the next year the program will be sufficiently reliable to accurately model the diverse effects of spacecraft charging.

Other critical aspects of the spacecraft charging phenomenon are the effects on contaminants and the effects of discharges. Since insulators tend to charge to different potentials than the satellite as a whole, it is possible to get preferential deposition of contaminant ions on such critical surfaces as optical sensors and solar cell covers (see Cauffman, 1973, for an analysis of contamination rates). Likewise arcing due to differential charging is apparently well documented (Figure 13). Although the actual arcing process itself is poorly known, fairly sophisticated models exist of the effects of the resultant electromagnetic pulse (Mindel, 1977).

As the emission of photoelectrons could be a potentially important source of contamination and has a significant impact on charging, a brief review is in order. Although the process of photoemission has been well understood



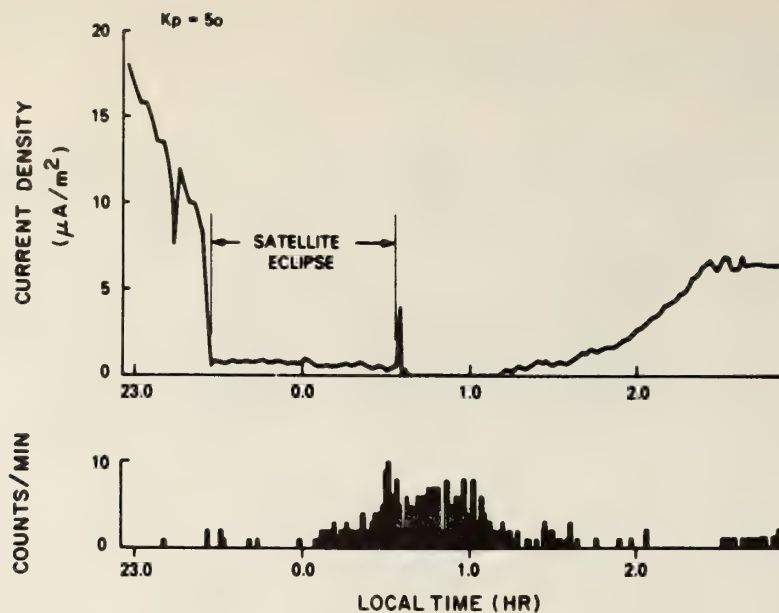


FIGURE 13. Arc-discharge rate and photoelectron current density as a function of local time during eclipse passage. The depression in the photoelectron current following eclipse passage is a result of spacecraft charging associated with an injection event. Thus the increased discharge rate is believed to be related to geomagnetic activity. From Shaw et al. (1976).

since Einstein's original paper, the actual emission rates for satellite materials are, in practice, not well known. Grard (1973) has made an extensive analysis of such phenomena. His results for common materials are presented in Table 1. Garrett and Rubin (1978) and Garrett and DeForest (1979a) have attempted by increasingly more sophisticated theories to predict the effects of a varying photoelectron flux. A typical result is given in Figure 14 demonstrating how the satellite potential can vary by 10 KV in a matter of a minute. As previously discussed, the escape rate (as opposed to emission rate) for photoelectrons appears to be  $.4 \text{ nA/cm}^2$ , a number larger on the average than that of the ambient flux of electrons at geosynchronous orbit (Garrett, 1977). The effect of such fluxes from large space structures on the environment has not been calculated. Satellites are known to generate wakes (Parker, 1977) which in turn could generate waves in the magnetospheric plasma. The large space structures currently planned could generate significant wave disturbances which might lead to instabilities in the low energy plasma environment. Parker (1977) and Douglas et al. (1977) have advanced simple models of this phenomenon. For example, in Douglas et al. (1977), the waves are envisioned as interacting with the structure itself (i. e., the solar sail) leading in some cases to resonance between the structure and the waves.

Another source of contamination is the addition of argon and other ionized gases from extensive space operations. A model of argon emission in the ionospheric domain advanced by Luhmann et al. (1978) has already been mentioned. Briefly,  $\sim 1.5 \times 10^{31} \text{ Ar}^+$  ions (compared with  $\sim 4 \times 10^{31}$  ambient

Material	Work Function (ev)	Electron Saturation Current ( $10^{12}/\text{sec m}^2$ )	Saturation Current Density ( $\mu\text{a/m}^2$ )
Aluminum Oxide	3.9	260	42
Indium Oxide	4.8	190	30
Gold	4.8	180	29
Stainless Steel	4.4	120	20
Aquadag	4.6	110	18
LiF on Au	4.4	90	15
Vitreous Carbon	4.8	80	13
Graphite	4.7	25	4

TABLE 1. Photoelectron emission characteristics of various spacecraft materials  
(Grard, 1973).



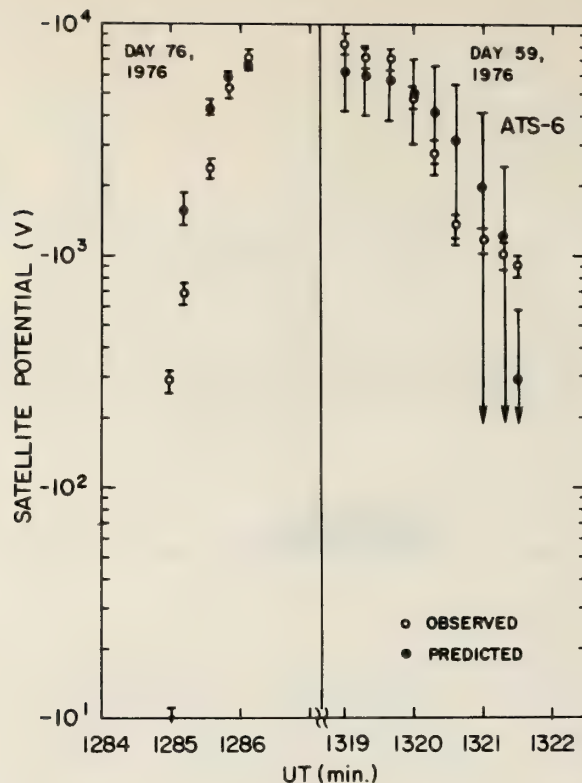


FIGURE 14. Observed and predicted potential variations on ATS-6 during passage through the earth's penumbra (Garrett and DeForest, 1979a).

ions about 500 km) are emitted as a solar power satellite slowly spirals out to geosynchronous orbit. The argon ions require a corresponding electron population for charge neutrality. Eventually, the heavy argon ions sink back into the ionosphere. As insufficient hydrogen ions exist to replace the argon, oxygen is drawn upwards giving the behavior illustrated in Figure 5.

Vondrak (1977, 1979) has presented a simple model of the leakage of neutral particles in the magnetosphere. Depending on the constituent, ionization rates on the order of hours to days could result in the slow buildup of a neutral gas ring followed, at equilibrium, by addition of charged particles to the plasmasheet flow. As Vondrak (1979) discusses, the total mass flow rate for the plasma sheet is only on the order of 1.2 kg/sec. The artificial addition of charged particles due to ion thrusters and/or a neutral gas ring (with subsequent ionization) could be of equal or higher magnitude to this flow rate. No detailed models for the effects on the auroral (plasmasheet) domain yet exist, however.

#### Magnetospheric Predictions

The preceding two sections have outlined the status of models of the plasma environment and of the various types of interactions. As the purpose of this conference is to predict, the review cannot be complete without also describing the relevant models for predicting the environmental parameters needed to

estimate the low energy particle interactions. These models fall into three areas - predictions of magnetospheric magnetic and electric field, direct predictions of ambient plasma parameters, and predictions of geomagnetic indices. The use of real time in situ data will be discussed in a later section.

Considering the availability and sophistication of magnetic field models (Olsen and Pfitzer, 1979), it is possible to make a reasonable approximation to these fields. Unfortunately, most magnetic field models are either for quiet periods or tied to  $K_p$  or  $a_p$  - earth-based indices. Likewise, the electric field models, such as that of Eq. 3, are all based on geomagnetic indices. Given the fields by any of these predictive methods or by actual in situ measurements, one could in principle predict the evolution of the plasma fluxes in time and space using the static or time-dependent models discussed previously. Once the fluxes are known, the interactions can be predicted.

Although the preceding technique is probably the most scientifically pleasing as it allows a consistent development of a prediction procedure for the ambient fluxes, it does not necessarily assure accurate predictions. The direct prediction of plasma parameters is to be preferred for modeling satellite interactions as the number of assumptions (i.e., particle drift, particle losses, etc.) are significantly reduced. Again, however, some means must be devised to determine the state of the magnetosphere. Even in the case of the simplest model, that of Chan et al. (1978), some input is required to tell how often to use the high or low profile. The models of Su and Konradi (1977) and Garrett and DeForest (1979b) both require knowledge of  $K_p$  or  $a_p$ . As yet, there is no way to predict these geosynchronous plasma parameters from interplanetary parameters.

The following indicates that no matter where one starts, the prediction of geomagnetic indices is the primary means of determining spacecraft interactions. In fact, as indicated in Figure 3 for the rate of arcing and in Figure 15 for the eclipse potential on ATS-5 and ATS-6, it may be possible to directly predict satellite interactions from ground-based indices. The prediction of geomagnetic indices of course depends primarily on persistence or solar wind parameters. A variety of studies have sought to correlate the various measured solar wind parameters, such as velocity, density, temperature, and magnetic field with geomagnetic activity (Schatten and Wilcox, 1967; Ballif et al., 1967, 1969; Hirshberg and Colburn, 1969; Arnoldy, 1971; Foster et al., 1971; Kane, 1972; Garrett, 1974, Garrett et al., 1974; Bobrov, 1973). These previous studies concentrated primarily on correlations with  $a_p$ ,  $K_p$ , Dst, and AE. An example of the adequacy of such models is illustrated in Figure 16 from Garrett et al., 1974. Most subsequent studies have essentially refined these early studies (see for example, Garrett et al., 1978a). Papers have also been presented at this conference (Saito, 1979; Clauer and McPherron, 1979; Iyemori and Maeda, 1979) on similar prediction schemes. It would be hoped that eventually a strong solar wind-magnetospheric plasma prediction scheme



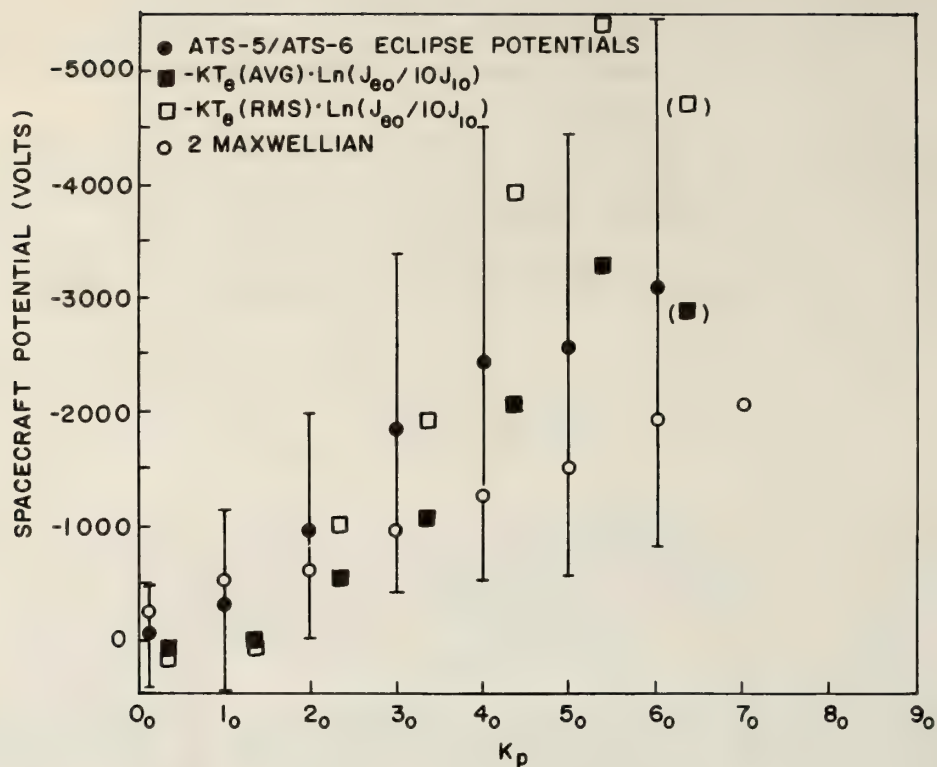


FIGURE 15. Observed and predicted potential variations for ATS-5 and ATS-6 during passage through the earth's shadow (Garrett et al., 1979a).

would also be found but as yet this has not happened.

The conclusion of all the preceding studies is that significant correlations do exist between measurable solar wind quantities and geomagnetic variations an hour to a few hours later. It still remains true, however, that persistence (i.e., if  $a_p$  is a given value now it will be the same in an hour) is a much better predictor in the time range of 0 to 10 hours than any of the solar wind quantities so far studied (Garrett et al., 1978a). Thus, in order to predict spacecraft interactions with any degree of confidence at present, actual ground-based measurements (in lieu of in-situ solar wind measurements) remain the most profitable prediction scheme. It remains to be determined, however, what the "best" geomagnetic parameter is for a given interaction process.

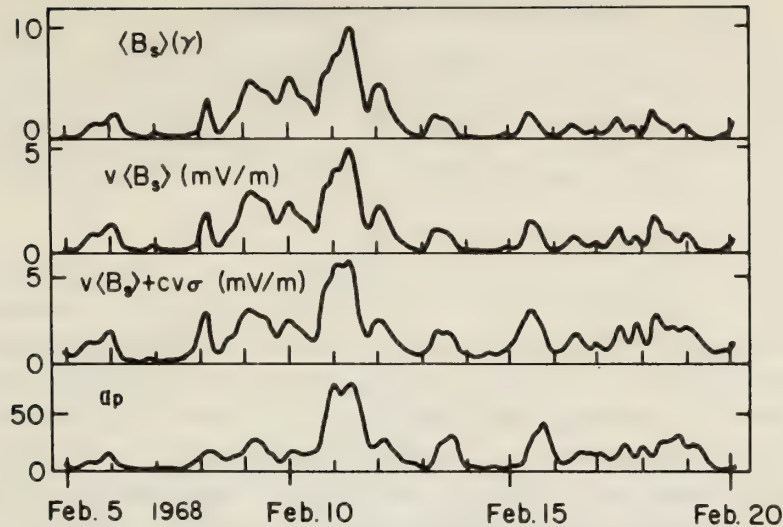


FIGURE 16. Relationship between various solar wind parameters and geomagnetic activity as measured by  $a_p$  (Garrett et al., 1974).

#### Representative Example

In another paper presented at this conference (Garrett et al., 1979a), an explicit example of the procedures outlined above has been worked out. Although it is not intended to repeat that paper here, the general procedures outlined in that paper demonstrate a typical problem and solution in predicting spacecraft interactions. From that standpoint, the findings of the paper will be briefly reviewed.

An important problem in spacecraft charging is the prediction of the extreme potential stresses likely to be encountered. As the largest measured potentials have been observed between the ATS-5 and ATS-6 geosynchronous satellites during eclipse passage, these data represent a useful starting point for estimating extreme potential effects (in reality the satellite to space potential during eclipse may well be an accurate reflection of the maximum differential potentials between two electrically isolated surfaces - one of which is shadowed and one which is illuminated). In Garrett et al. (1979a), the eclipse potentials for seven eclipse seasons of ATS-5 and ATS-6 were carefully measured and plotted as a function of  $K_p$ . These results are plotted in Figure 15 as closed dots with error bars. This plot obviously serves as a simple predictive model.

Two charging theories were advanced by Garrett et al. (1979a) to explain the charging phenomenon based on ambient plasma parameters. The simplest being based on the formula

$$/q V_o/ \cong - K T_e \ln (J_{E0}/10 J_{I0}) \quad (5)$$



where:  $q$  = electronic charge  
 $V_o$  = satellite potential in volts  
 $T_e$  = electron temperature  
 $J_{E0}$  = ambient electron current  
 $J_{I0}$  = ambient ion current  
 $K$  = Boltzmann's constant

Likewise, two models of the environment were advanced. One being based on statistical tabulations of  $T_e$ ,  $J_{E0}$ , and  $J_{I0}$  versus  $K_p$  and the other based on the variation of the distribution function at  $\sim 0130$  local time with  $K_p$  (based on the model described in Eq. 2 and plotted in Figure 9). The two environmental models, both keyed to  $K_p$ , were then combined with the potential prediction codes to estimate the potential as a function of  $K_p$ . The results are plotted in Figure 15 and represent good agreement between theory and observation.

## FUTURE RESEARCH

Considering the potential importance of models of interactions between spacecraft and the environment, it is a significant fact that only in the last two or three years have attempts been made to accurately define the ambient low-energy plasma environment. Further, it has been only since the recent launch of the SCATHA satellite that any significant attempt has been made to characterize the interaction processes. Thus, it is not difficult to find many inadequacies in the present state of spacecraft/environment interaction prediction.

Surprisingly, much data exists currently on the low-energy near-earth environment - data which is yet to be exploited to any great degree. The geosynchronous orbit is on the whole quite adequately covered and, with instruments capable of measuring ionic composition and field aligned fluxes on GEOS I, GEOS II, and SCATHA, this region should soon be well analyzed. The problem is still, however, how best to characterize this rapidly varying regime for real-time predictions. In Garrett *et al.* (1979b), data from ATS-6 and the geosynchronous satellite 1976-059A are compared. The 1976-059A returns in real time the electron fluxes from 30 keV up (Higbie *et al.*, 1979). This information appears adequate for predicting the current state of the magnetosphere (Higbie *et al.*, 1979; Baker *et al.*, 1979) to the degree necessary to predict spacecraft charging. Thus, such real time measurements may offer one possible solution to the prediction problem for geosynchronous orbit (see also Thompson and Secan, 1979). The issue for the plasmasphere and near-earth high latitude region is not as clear cut. Although sufficient data apparently exists for the equatorial plasmasphere to construct meaningful statistical models, developments in this area have not been forthcoming as yet. Recently, a tremendous amount of data for detailed studies of the low energy particle fluxes at high latitudes has become available from the S3-2,

S3-3, and DMSP/F2 (J/3 sensor) satellites (Burke et al., 1979). It is hoped that this data base will soon make possible significant advances in this region also. The DMSP photos may ultimately provide a real time method of evaluating rapid variations in this region as 1976-059A may for geosynchronous orbit. Likewise, ISEE-3 and OPEN-A may allow real-time solar wind monitoring (Cauffman, 1979).

From a user standpoint it is of critical importance that accurate statistical models be available in the future if any long term mission planning is to be carried out. Likewise, the user will need adequate models of the interaction phenomena involved. Further, little progress can be made in the future in predicting for the user spacecraft/environment interactions if the required parameters are unknown. It does little good to provide detailed information on the energy flux if the interaction process is dependent on number flux. Unfortunately, the biggest gap in this area is in material response functions. This is not a problem of this conference but until it is resolved, the only part of the spacecraft/interactions problem that we can accurately attack is the ambient space environment and its perturbations.

Predicated on the following it appears that future research should be directed in main areas:

1. Defining in a statistical sense the ambient environment based on user requirements (particularly the near-earth plasmasphere and at high latitudes).
2. Deriving analytical expressions or efficient 3-dimensional models (static and time-dependent) capable of simulating magnetospheric variations for input into interaction modeling programs.
3. As it currently appears that predictions based on earth-based or solar wind parameters will not approach the accuracy of in situ measurements, expanded real-time in situ measurements must be made.
4. Verification of particle drift theory so that static (or time-dependent) models can be used to extend measurements made in limited orbital regions to other regions.
5. Refinement of our basic understanding of interaction phenomena.

In the next few years, major steps will be made in all these areas. Active experiments such as barium releases and joint SCATHA/GEOS beam studies should greatly improve our ability to trace contaminant clouds. Likewise, the Long Duration Exposure Facility (LDEF) should provide important information on the long term effects of the environment. Support of the space physics community should be given to projects such as these as they promise immediate, practical returns from our scientific studies.



## CONCLUSION

In this review the present state of spacecraft/environment interactions modeling and prediction as it pertains to the low energy plasma environment has been covered. A multitude of issues have been raised in each of the main areas - modeling the interaction process, modeling the environment, and predicting the future state of the magnetosphere (and, hence, the magnitude of the interaction). As should be apparent, the status of the modeling efforts are still somewhat rudimentary - the most success having been in developing static models of the magnetosphere, the least in determining material responses. The role of the user has had little to do in determining the research trends in this area as it has only comparatively recently been realized that spacecraft/environment interactions were of real importance. As more emphasis is placed on user needs, however, we should expect a shift to more basic models of the environment (i.e., statistical models like those of the radiation environment). Ultimately, however, the 3-dimensional, time dependent models will be necessary if a complete understanding is desired. In the interim, in situ, real-time monitoring appears to be the only plausible solution.

## REFERENCES

- Angerami, J. J., and J. O. Thomas (1964): Studies of planetary atmospheres, 1. The distribution of electrons and ions in the earth's exosphere. J. Geophys. Res., 69, 4537-4559.
- Arnoldy, R. L. (1971): Signature in the interplanetary medium for substorms. J. Geophys. Res., 76, 5189-5201.
- Baker, D. N., P. R. Higbie, E. W. Hones, and R. D. Belian (1979): The use of 30 keV electron anisotropies at  $6.6 R_E$  to predict magnetospheric substorms. To be published in the Proceedings of the International Solar-Terrestrial Predictions Workshop.
- Ballif, J. R., D. E. Jones, P. J. Coleman, L. Davis, and E. J. Smith (1967): Transverse fluctuations in the interplanetary magnetic field: A requisite for geomagnetic variability. J. Geophys. Res., 72, 4357-4364.
- Ballif, J. R., D. E. Jones, and P. J. Coleman (1969): Further evidence of the correlation between transverse fluctuations in the interplanetary magnetic field and  $K_p$ . J. Geophys. Res., 74, 2289-2301.
- Bobrov, M. S. (1973):  $K_p$  index correlations with solar-wind parameters during the first and second stages of a recurrent geomagnetic storm. Planet. Space Sci., 21, 2139-2147.
- Brundin, C. L. (1963): Effects of charged particles on the motion of an earth satellite. AIAA J., 1, 2529-2538.
- Burke, W. J., D. A. Hardy, F. J. Rich, M. C. Kelley, M. Smiddy, B. Shuman, R. C. Sagalyn, R. P. Vancour, P. J. L. Wildman, S. T. Lai, and J. Bass (1979): A case study of S3-2 observations in the late evening auroral oval. To appear in AFGL-TR series report.
- Carpenter, D. L. (1966): Whistler studies of the plasmapause in the magnetosphere. 1. Temporal variations in the position of the knee and some evidence on plasma motions near the knee. J. Geophys. Res., 71, 693-709.
- Carpenter, D. L., and C. G. Park (1973): On what ionospheric workers should know about the plasmapause-plasmasphere. Rev. Geophys., 11, p. 133-154.
- Cauffman, D. P. (1973): Ionization and attraction of neutral molecules to a charged spacecraft. SAMSO-TR-73-263.



- Caufmann, D. P. (1979): Space plasma monitoring: Two new capabilities for the coming decade. To be published in the Proceedings of the International Solar Terrestrial Predictions Workshop.
- Chan, K. W., D. M. Sawyer, and J. I. Vette (1977): A model of the near-earth plasma environment and application to the ISEE-A and -B orbit. NSSDC/WDC-A R&D 77-01.
- Chappell, C. R., K. K. Harris, and G. W. Sharp. (1970): The morphology of the bulge of the plasmasphere. J. Geophys. Res., 75, 3848.
- Chappell, C. R. (1972): Recent satellite measurements of the morphology and dynamics of the plasmasphere. Rev. Geophys., 10, 951-979.
- Chen, A. J. (1970): Penetration of low-energy protons deep into the magnetosphere. J. Geophys. Res., 75, 2458-2467.
- Chen, A. J., and R. A. Wolf (1972): Effects on the plasmasphere of a time-varying convection electric field. Planet. Space Sci., 20, 483-509.
- Chiu, Y. T., J. G. Luhmann, B. K. Ching, and D. J. Boucher, Jr. (1978): An equilibrium model of plasmaspheric composition and density. Submitted to J. Geophys. Res.
- Chiu, Y. T., J. G. Luhmann, B. K. Ching, M. Shulz, and D. J. Boucher (1979): Magnetospheric and ionospheric impact of large-scale space transportation with ion engines. SAMSO-TR-79-3.
- Clauer, C. R., and R. L. McPherron (1979): Predicting partial ring current development. To be published in the Proceedings of the International Solar Terrestrial Predictions Workshop.
- DeForest, S. E., and C. E. McIlwain (1971): Plasma clouds in the magnetosphere. J. Geophys. Res., 76, 3587-3611.
- DeForest, S. E. (1972): Spacecraft charging at synchronous orbit. J. Geophys. Res., 77, 651.
- DeForest, S. E., and A. R. Wilson (1976): A preliminary specification of the geosynchronous plasma environment. DNA 3951T.
- Douglas, M., R. LaQuey and S. DeForest (1977): The effect of environmental plasma interactions on the performance of the solar sail system. NASA CR-135258.

- Ejiri, M., R. A. Hoffman and P. H. Smith (1977): The convection electric field model for the magnetosphere based on Explorer 45 observations. NASA/GSFC, X-625-77-108. To appear in J. Geophys. Res., 1978.
- Foster, J. C., D. H. Fairfield, K. W. Ogilvie, and T. J. Rosenberg (1971): Relationship of interplanetary parameters and occurrence of magnetospheric substorms. J. Geophys. Res., 76, 6971-6975.
- Frank, L. A. (1967): On the extraterrestrial ring current during geomagnetic storms. J. Geophys. Res., 72, 3753-3767.
- Freeman, J. W. (1974):  $K_p$  dependence of the plasma sheet boundary. J. Geophys. Res., 79, 4315.
- Freeman, J. W., and P. H. Reiff (1979): Environmental protection for the solar power satellite. To appear in Proceedings of the Spacecraft Charging Technology Conference-II (edited by R. R. Lovell and C. P. Pike).
- Garrett, H. B. (1974): The role of fluctuations in the interplanetary magnetic field in determining the magnitude of substorm activity. Planet. Space Sci., 22, 111-119.
- Garrett, H. B., A. J. Dessler, and T. W. Hill (1974): Influence of solar wind variability on geomagnetic activity. J. Geophys. Res., 79, 4603-4610.
- Garrett, H. B. (1977): Modeling of the geosynchronous orbit plasma environment - Part I. AFGL-TR-77-0288.
- Garrett, H. B. and A. G. Rubin (1978): Spacecraft charging at geosynchronous orbit - generalized solutions for eclipse passage. Geo. Res. Ltrs., 5, 865-868.
- Garrett, H. B., D. F. Smart, and M. A. Shea (1978a): A study of the use of interplanetary magnetic field and plasma measurements as a predictor of geomagnetic activity. Proceedings of the Special Committee on Solar-Terrestrial Physics, 1978.
- Garrett, H. B., E. G. Mullen, E. Ziemba, and S. E. DeForest (1978b): Modeling of the geosynchronous plasma environment - Part 2. ATS-5 and ATS-6 statistical atlas. AFGL-TR-78-0304.
- Garrett, H. B. (1979): Review of quantitative models of the 0 to 100 keV near-earth plasma. To appear in Rev. Geophys.



- Garrett, H. B., and S. E. DeForest (1979a): Effects of a time-varying photoelectron flux on spacecraft potential. To appear in J. Geophys. Res.
- Garrett, H. B., and S. E. DeForest (1979b): An analytical simulation of the geosynchronous plasma environment. To appear in Planet. Space Sci.
- Garrett, H. B., A. G. Rubin, and C. P. Pike (1979a): Prediction of spacecraft potentials at geosynchronous orbit. To appear in Solar-Terrestrial Prediction Proceedings.
- Garrett, H. B., D. Schwank, P. R. Higbie, and D. N. Baker (1979b): Comparison between 30-80 keV electron channels on ATS-6 and 1976-059A. Abstract to appear in EOS.
- Gauntt, D. M. (1979): Differential charging of large space structures. To appear as AFGL Tech Rept.
- Grard, R. J. L. (1973): Properties of the satellite photoelectron sheath derived from photoemission laboratory measurements. J. Geophys. Res., 78, 2885-2906.
- Grebowsky, J. M., and A. J. Chen (1975): Effects of convection electric field on the distribution of ring current type protons. Planet. Space Sci., 23, 1045.
- Hagler, T., H. G. Patterson, and C. A. Nathan (1977): Learning to build large structures in space. Aeronautics and Astronautics, 12, 51-57.
- Harel, M., R. A. Wolf, and P. H. Reiff (1978): Preliminary report of the first computer run simulating the substorm-type event of 19 September 1976<sup>1</sup>. Appendix A, Annual NSF Report for Grants ATM 74-21185 and ATM 74-21185A01.
- Higbie, P. R., D. N. Baker, R. D. Belian, and E. W. Hones, Jr., (1979): Evolution of substorm and quiet-time electron anisotropes ( $30 \leq E_e \leq 300$  keV). To appear in Solar-Terrestrial Prediction Proceedings.
- Hirshberg, J., and D. S. Colburn (1969): Interplanetary field and geomagnetic variations - a unified view. Planet. Space Sci., 17, 1183-1206.
- Inouye, G. T. (1976): Spacecraft potentials in a substorm environment. Spacecraft Charging by Magnetospheric Plasma, AIAA Progress in Astronautics and Aeronautics, Series 42. (edited by A. Rosen).

- Iyemori, T., and H. Maeda (1979): Prediction of geomagnetic activities from solar wind parameters based on linear prediction theory. To be published in the Proceedings of the International Solar-Terrestrial Workshop.
- Jaggi, R. K., and R. A. Wolf (1973): Self-consistent calculation of the motion of a sheet of ions in the magnetosphere. J. Geophys. Res., 78, 2852-2866.
- Johnson, W. G., T. A. Parnell, W. W. Vaughn (1979): Solar-induced variations of environment affecting manned space flight and spacecraft operations. To be published in the Proceedings of the International Solar-Terrestrial Predictions Workshop.
- Kane, R. P. (1972): Relationship between the various indices of geomagnetic activity and the interplanetary plasma parameters. J. Atmos. Terr. Phys., 34, 1941-1943.
- Katz, I., M. J. Mandell, G. W. Schnuelle, J. J. Cassidy, and J. C. Roche (1979): The capabilities of the NASA charging analyzer program. To appear in Proceedings of the Spacecraft Charging Technology Conference-II (edited by R. K. Lovell and C. P. Pike).
- Kavanagh, L. D., Jr., J. W. Freeman, Jr. and A. J. Chen (1968): Plasma flow in the magnetosphere. J. Geophys. Res., 73, 5511.
- Kivelson, M. G., S. M. Kaye, and D. J. Southwood (1978): The physics of plasma injection events. Presented at AGU Chapman Conference on Magnetospheric Substorms, UCLA Institute of Geophysics and Planetary Physics, Pub. No. 1878.
- Konradi, A., C. L. Semar, and T. A. Fritz (1975): Substorm-injected protons and electrons and the injection boundary model. J. Geophys. Res., 80, 543-552.
- Laframboise, J. A., and S. M. L. Prokopenko (1977): Numerical simulation of spacecraft charging phenomena. Proc. of the Spacecraft Charging Conference, AFGL-TR-77-0051/NAS A-TMX-73537 (edited by C. P. Pike and R. K. Lovell) 309-318.
- Lemaire, J., and M. Scherer (1974): Exospheric models of the topside ionosphere. Space Sci. Rev., 15, 591.
- Lennartsson, W., and D. L. Reasoner (1978): Low-energy plasma observations at synchronous orbit. J. Geophys. Res., 83, p. 2145-2156.



- Luhmann, J. G., Y. T. Chin, B. K. Ching (1978): Argon ion propellants in the plasmasphere. Presented at The Solar-Terrestrial Physics Symposium, Innsbruck, 207-211.
- Massaro, M. J., T. Green, and D. Ling (1977): A charging model for three-axis stabilized spacecraft. Proceedings of the Spacecraft Charging Technology Conference (edited by C.P. Pike and R.R. Lovell), AFGL-TR-77-0051/NASA TMX-73537, 237-267.
- Mauk, B. H. and C. E. McIlwain (1974): Correlation of  $K_p$  with the substorm plasma sheet boundary. J. Geophys. Res., 79, 3193-3196.
- Mayr, H. G., E. G. Fontheim, L. H. Bruce, H. C. Brinton, and H. A. Taylor, Jr. (1972): A. Theoretical model of the ionosphere dynamics with interhemispheric coupling. J. Atmos. Terr. Phys., 34, 1659.
- McCoy, J., A. Konradi, and O. K. Garriott (1979): Current leakage for low altitude satellites. AIAA Paper 79-0387, 17th Aerospace Sciences Mtg.
- McIlwain, C. E. (1972): Plasma convection in the vicinity of the geosynchronous orbit. Earth's Magnetospheric Processes, edited by B. M. McCormac, p. 268, D. Reidel, Dordrecht, Netherlands.
- McPherson, D. A., and W. R. Schober (1976): Spacecraft charging at high altitudes: the SCATHA satellite program. Spacecraft Charging by Magnetospheric Plasmas (edited by A. Rosen), AIAA Progress in Astronautics and Aeronautics, 47, 15-30.
- Mindel, I. N. (1977): DNA EMP Awareness Course Notes, 3rd Ed., DNA 2772T.
- Olson, W. P. and K. A. Pfizter (1979): Magnetospheric magnetic and electric field models and their role in spacecraft interaction modeling. AIAA Paper 79-0388, 17th Aerospace Sciences Meeting.
- Parker, L. W. (1977): Calculation of sheath and wake structure about a pill-box-shaped spacecraft in a flowing plasma. Proceedings of the Spacecraft Charging Technology Conference (edited by C.P. Pike and R.R. Lovell), AFGL-TR-77-0051/NASA TMX-73537, 331-366.
- Roederer, J. G., and E. W. Hones, Jr. (1970): Electric field in the magnetosphere as deduced from asymmetries in the trapped particle flux. J. Geophys. Res., 75, 3923-3926.
- Roederer, J. G. and E. W. Hones, Jr. (1974): Motion of magnetospheric particle clouds in a time-dependent electric field model. J. Geophys. Res., 1432-1438.

- Rothwell, P. L., Rubin, A. G., and Yates, G. K. (1977): A simulation model of time-dependent plasma-spacecraft interactions. Proc. of the Spacecraft Charging Conf., AFGL-TR-77-0051/NASA TMX-73537 (Edited by C. Pike and R. R. Lovell) 389-412.
- Saito, T. (1979): Prediction of substorm activities. To be published in the Proceedings of the International Solar-Terrestrial Predictions Workshop.
- Schatten, K. H. and Wilcox, J. M. (1967): Response of the geomagnetic index  $K_p$  to the interplanetary magnetic field. J. Geophys. Res., 72, 5185-5192.
- Schnuelle, G. W., D. E. Parks, I. Katz, M. J. Mandell, P. G. Steen, J. J. Cassidy and A. Rubin (1979): Charging analysis of the SCATHA satellite. To appear in Proceedings of the Spacecraft Charging Technology Conference (edited by R. R. Lovell and C. P. Pike).
- Schunk, R. and J. C. G. Walker (1969): Thermal diffusion in the topside ionosphere for mixtures which include multiply charged ions. Planet. Space Sci., 17, 853.
- Shaw, R. R., J. E. Nanevich and R. C. Adamo (1976): Observations of electrical discharges caused by differential satellite-charging. Spacecraft Charging by Magnetospheric Plasmas (edited by A. Rosen), AIAA Progress in Astronautics and Aeronautics, 47, 61-76.
- Smith, P. H., N. K. Beutra and R. A. Hoffman (1978a): Inference of the ring current ion composition by means of charge exchange decay. NASA TM-79611.
- Smith, P. H., R. A. Hoffman and N. K. Beutra (1978b): A visual description of the dynamical nature of magnetospheric particle convection in a time-varying electric field. Abstract in Trans. Am. Geophys. Union, 59, 361.
- Smith, R. H., N. K. Beutra and R. A. Hoffman (1979): Motions of charged particles in the magnetosphere under the influence of a time-varying large scale convection electric field: Quantitative Modeling of the Magnetospheric Processes, Geophys. Monogr. Ser. 21, edited by W. P. Olson, AGU, Washington, D. C.
- Southwood, D. J. (1977): The role of hot plasma in magnetospheric convection. J. Geophys. Res., 35, 5512-5520.



- Su, S.-Y. and A. Konradi (1977): Description of the plasma environment at geosynchronous altitude. NASA Johnson Technical Note, p-10.
- Thompson, R. L., and J. A. Secan (1979): Geophysical forecasting at AFGWC. To be published in the Proceedings of the International Solar-Terrestrial Predictions Workshop.
- Vasyliunas, V. M. (1968): A survey of low-energy electrons in the evening sector of the magnetosphere with OGO 1 and OGO 3. J. Geophys. Res., 73, 2839-2884.
- Vasyliunas, V. M. (1972): Magnetospheric plasma. Solar Terrestrial Physics, 1970. Proc. of the International Symposium, Leningrad, USSR, May 11-19, 1970, edited by E. R. Dyer, D. Reidel, Dordrecht, Netherlands.
- Vondrak, R. R. (1977): Environmental impact of space manufacturing. Paper 77-539, Third Princeton/AIAA Conference on Space Manufacturing Facilities.
- Vondrak, R. R. (1979): The effects of gas releases from large space structures. AIAA Paper 79-0390, 17th Aerospace Sciences Mtg.
- Walker, R. J., and M. G. Kivelson (1975): Energization of electrons at synchronous orbit by substorm-associated cross-magnetospheric electric fields. J. Geophys. Res., 80, No. 16, p. 2071-2082.
- Whipple, E. C. (1965): The equilibrium electric potential of a body in the upper atmosphere. NASA X-615-65-296.
- Whipple, E. C., and L. W. Parker (1969a): Theory of an electron trap on a charged spacecraft. J. Geophys. Res., 74, 2962-2971.
- Whipple, E. C., and L. W. Parker (1969b): Effects of secondary electron emission on electron trap measurements in the magnetosphere and solar wind. J. Geophys. Res., 74, 5763-5774.
- Wolf, R. A. (1970): Effects of ionospheric conductivity on convective flow of plasma in the magnetosphere. J. Geophys. Res., 75, 4677-4698.
- Wolf, R. A. (1974): Calculation of magnetospheric electric fields. Magnetospheric Physics, edited by B. M. McCormac, D. Reidel, Dordrecht, Netherlands, 167-177.
- Young, D. T. (1979): Ion composition measurements in magnetospheric modeling. Quantitative Modeling of the Magnetospheric Processes, Geophys. Monogr. Ser., 21, edited by W.P. Olson, AGU, Washington, DC.

# PREDICTABILITY OF UPPER-ATMOSPHERIC DENSITY AND COMPOSITION

Joe M. Straus and David R. Hickman  
Space Sciences Laboratory  
Ivan A. Getting Laboratories  
The Aerospace Corporation  
P.O. Box 92957  
Los Angeles, California 90009

Empirical models of upper-atmospheric density and composition are employed in a number of areas, ranging from basic research in atmospheric and ionospheric physics to practical applications in satellite ephemeris prediction. Such models have been based on various kinds of data sets and have varying levels of complexity, strengths, and weaknesses. This paper describes the characteristics of several of the widely used models and reviews studies in which the predictions of these models have been compared with observational data. A final section presents a discussion of the relative advantages and limitations of the models in current use, as well as a brief discussion of ways in which the models might be improved.

## 1. DESCRIPTION OF UPPER-ATMOSPHERIC MODELS

The development of atmospheric models has a history dating back at least a hundred years. At thermospheric altitudes, however, there were few measurements prior to the launch of Sputnik in 1957. The resulting lack of knowledge of the details of atmospheric behavior made it impossible for models to have any predictive capability. Therefore, models developed prior to 1957 were used primarily as standards, allowing a common basis for calculations performed by diverse investigators.

ARDC 1959 (Minzner et al., 1959) was the first standard upper atmospheric model that utilized satellite measurements of the density. Densities were deduced by the orbital decay method, and so only very coarse time and space resolution was provided by the measurements. The development of standard atmospheres has continued, leading to the U.S. Standard Atmosphere, 1962 (COESA, 1962) and the U.S. Standard Atmosphere, 1976 (COESA, 1976). Such standard atmospheres are not intended to have significant predictive uses and will not be further discussed here. A recent review of these standard atmospheres has been given by Minzner (1977).

L. G. Jacchia has, over the years, developed a series of thermospheric models. His first (Jacchia, 1960) was based on drag data from four satellites



having perigee altitudes between 210 and 650 km. The model took into account both diurnal effects and variations related to solar activity. Geomagnetic activity effects had been observed in orbital decay data, but were not sufficiently well understood to be modeled at that time. Unlike the ARDC 1959 model, which was presented in tabular form, the Jacchia (1960) model density was expressed analytically by

$$\rho = \rho_0 \frac{F_{20}}{10} \{ 1 + 0.19 [\exp (0.0055 z) - 1.9] \cos^6 \frac{\psi}{2} \}$$

where  $\log \rho_0 = -16.021 - 0.001985 z + 6.363 \exp (-0.0026 z)$ . In this expression,  $\rho$  is the density in  $\text{gm/cm}^3$  at altitude  $z$  in km ( $200 \leq z \leq 700$ );  $F_{20}$  is the 20-cm solar flux in units of  $10^{-22} \text{ W/m}^2\text{-Hz}$ ; and  $\psi$  represents the angular distance from the maximum of the diurnal density bulge, which was placed at the subsolar latitude at a local time of about 14 hr. [Note that Jacchia (1960) used the 20-cm solar flux as an index of the degree of solar activity. Subsequent studies utilized the 10.7-cm flux which was found to be somewhat better correlated with density variations.] The simplicity of the model made it particularly useful for production-type calculations (as in orbit determination) where numerical efficiency is required.

The CIRA-1961 (COSPAR International Reference Atmosphere) model consisted of three tables of the atmospheric variables that represented the daily average conditions, daytime maximum, and nighttime minimum conditions. No other time or space variation of the atmosphere was described. The model purported to be representative of medium solar activity conditions, but since the data base consisted of orbital decay observations in the 1958-60 period, it should be regarded as a high solar activity model. The model has almost no predictive capability and was of importance primarily because of its internationally recognized status.

The work of Harris and Priester (1962a, b) represents a milestone in thermospheric modeling studies since theirs was a first major attempt to describe the upper atmosphere on a physical basis. Rather than fit the observational data with profiles or empirical formulas, they fit the data subject to the constraints of the thermospheric energy equation under conditions of hydrostatic equilibrium. In other words, they solved the one-dimensional time-dependent heat condition equation to obtain the temperature, and with this the different constituent profiles could be computed assuming diffusive equilibrium. Once the temperature and the concentrations of the different species are known, all other atmospheric variables are determined. To fit the computed density to the observations, the boundary conditions or properties of the energy sources can be adjusted.

Harris and Priester encountered some difficulty in trying to match the observations using only the solar EUV flux as the source of heating. Using a second heat source to supplement EUV heating, which was later shown to be needed because of advective heat transport due to thermospheric winds, Harris and Priester (1962a) produced a diurnal model of the upper atmosphere that was appropriate for equatorial and temperate latitudes at equinox and for solar activity conditions corresponding to  $F_{10.7} = 200$ . Shortly thereafter, Harris and Priester (1962b) extended their model to include solar cycle variations. COSPAR (Congress on Space Research) adopted the Harris-Priester models for the CIRA-1965 model. Under this title, 120 tables were published that described the thermospheric structure from 120-800 km for 10 levels of solar activity ( $65 < F_{10.7} < 225$ ) and with 2-hour increments of local time.

The tables, as described above, were based on theory. However, to account for other observed density effects, such as the 27-day variation, geomagnetic activity effects and semiannual changes, empirically derived corrections to the exospheric temperature need to be used in conjunction with the tables.

With satellite drag data from half a solar cycle (1958-1964), Jacchia (1965) was able to construct an improved empirical model that took into account solar activity (including both 11-year and 27-day cycles), diurnal and semiannual variations, and geomagnetic activity. Although the published model was accompanied by tables of the atmospheric quantities, the composition, temperature, and density could be directly computed for any given set of conditions. This is in contrast to the CIRA-1965 model, in which interpolation in the tables was required for arbitrary conditions. The formulation of the Jacchia (1965) model was also continuous over the globe, whereas CIRA-1965 was essentially only for middle and low latitudes.

With this model Jacchia initiated the now common practice of using the exospheric temperature as a parameter to synthesize the density data at different altitudes. To elaborate, given a lower boundary temperature and the exospheric temperature, a temperature profile is assumed to be determined. Then, given a set of lower boundary conditions for number densities of the constituents, the equations for diffusive equilibrium are integrated to determine the vertical distributions of the different atmospheric constituents, which may be summed to obtain the total density:

$$\frac{dn_i}{n_i} = -\frac{m_i g}{kT(z)} dz - (1 + \alpha_i) \frac{dT}{T}$$

$$\rho = \sum_i n_i m_i$$

where  $n_i$  is the concentration of the  $i$ th constituent,  $m_i$  is its mass,  $g$  is the acceleration of gravity,  $z$  is the altitude,  $k$  is Boltzmann's constant,  $T$  is the temperature,  $\alpha_i$  is the thermal diffusion coefficient and  $\rho$  is the total mass density. In essence, then, given a set of lower boundary conditions, a density profile is determined by specifying the exospheric temperature. Thus, rather than manipulate the density to simulate the different atmospheric variations, Jacchia (1965) utilizes exospheric temperature changes to produce the density variations. The appropriate temperature changes, which were analytically formulated, were derived by fitting the model to observations.

In order to meet the requirements for a more realistic atmosphere than the U.S. Standard 1962, a model based on Jacchia (1965) was adopted by the U.S. Committee on Extension to the Standard Atmosphere (COESA). The description of this model was published as the U.S. Standard Atmosphere Supplements, 1966 (COESA, 1966).

By 1970 there had been a number of upper atmospheric composition measurements made by mass spectrometer and EUV absorption techniques. The results of these measurements indicated a need to revise the  $n(O)/n(O_2)$  ratio near 150 km, and a new model was produced by Jacchia (1970) to account for this. Following a review by von Zahn (1970) of density measurements near 150 km, Jacchia's 1970 model was quickly supplanted by a new (Jacchia, 1971) model which incorporated von Zahn's recommended density values at 150 km. This new model gained wide acceptance and use and was adopted virtually unchanged as CIRA-1972.



The CIRA-1972 model recognizes several types of variations which are dealt with explicitly. Specifically they are: solar cycle variations and shorter-term solar activity effects; the diurnal variation; geomagnetic activity variations; the semiannual variation; and seasonal-latitudinal variations. [More rapid density variations (waves) were also recognized as being present, but were regarded as unpredictable in a static model of this type.] The first two of the five variations listed above were accounted for using a variation of the exospheric temperature with the 10.7 cm solar flux and with local time and latitude. The last three variations were incorporated as direct percentage variations of the density.

CIRA-1972 represented the zenith of a long tradition in thermospheric models in which a fixed number density for each constituent was defined at a lower boundary and then the vertical profile for each constituent was determined from the temperature profile by means of a diffusive equilibrium equation. Already it had become necessary to adjust the resulting densities somewhat in order to agree with observations, but the basic foundation of the method was intact.

Although some direct measurements of atmospheric composition were used in the formulation of the CIRA-1972 model, the primary data source was satellite drag data. These data were further restricted to altitudes above ~200 km, leading to a model that represents total density rather well above that altitude. However, as will be discussed later, substantial discrepancies appear when this model is used to predict the composition of the upper atmosphere. As will be described shortly, analysis of densities measured by the mass spectrometer on the OGO-6 satellite revealed that the diurnal, seasonal, and geomagnetic variations of the different constituents are different and that the conceptually attractive picture of density variations being related to a temperature profile through a diffusive equilibrium equation is not adequate. In order to account for the (seemingly) unrelated variations of the individual constituents, Jacchia produced a new model (Jacchia, 1977) in which a separate "pseudo temperature" profile was introduced for each constituent. A number of additional corrections to a diffusive equilibrium type of behavior were also required. The resulting model has little of the conceptual attractiveness of the older models from which it descended, and, in addition, its great complexity makes its use awkward. As a result it has not been a widely used model.

An entirely new and powerful approach was developed by Hedin and his co-workers for producing a model based on OGO-6 mass spectrometer measurements. Von Zahn and his co-workers have used the same technique to develop a model based on ESRO-4 mass spectrometer data. Since these models are identical in their mathematical formulations, differing only in their data bases, they will be described together. In both models, the composition is described in terms of density variations at a reference altitude  $z_0$  (450 km for OGO-6 and 275 km for ESRO-4) with an altitude dependence which assumes isothermal conditions at higher altitudes at any given time and location. A spherical harmonic expansion (in geographic latitude-local time coordinates) is used to fit the density data at  $z_0$  (in a least-squares sense); the coefficients of this expansion were determined using the mass spectrometer data and are tabulated in the model report. A total of 37 coefficients is involved for each thermospheric constituent, and variations in density due to solar activity (F10.7), geomagnetic activity ( $A_p$ ), time of year (including annual and semi-annual terms), latitude and local time (including diurnal, semidiurnal, and terdiurnal terms) are treated. The form of the dependence of density on

F10.7 is similar to that of the Jacchia (1971) model, except that allowance is made for a possible non-linearity in the correlation with the 27-day F10.7 variation; such an effect has been suggested by some data from ground-based observations. The dependence of density on  $A_p$  included in these models is small; the possibility of a latitude variation of the geomagnetic activity effect has been included. Once the density distributions at the reference altitude and  $T_\infty$  are determined in this way, formulas based on hydrostatic equilibrium are used to evaluate the densities at all other altitudes above 120 km, where variations in all quantities other than the temperature and the  $N_2$  density are allowed.

The accuracy of these models is on the order of the experimental error for He and O (20% for O and larger for He) and about three times the experimental error for  $N_2$ . The root-mean-square error of the models (with respect to the data that were used in their generation) is on the order of 15-20%.

The merit of a model based on a least-squares fit to the data using spherical harmonic analysis is that these functions form a complete set that can, in principle, represent any degree of complexity in the data by an increase in the number of terms used. Furthermore, since the spherical harmonics are approximate eigenfunctions of the thermosphere, relatively few terms should be required. Using this approach, the individual behavior of the different constituents can be accommodated. Furthermore, as opposed to earlier models which include only those variations specifically noted by analysts and purposely incorporated in the model, this type of model can faithfully reflect atmospheric behavior present in that data, but not yet discovered by scientists.

There are two aspects of the mass spectrometer models that limit their usefulness. Since each model used data from only about two years of satellite operation, the coverage in latitude and local time at specified values of F10.7 and  $A_p$  is not really adequate for producing a definitive model. (This is especially evident with respect to  $A_p$ , leading to a restriction of the models to quiet geomagnetic conditions.) The second limitation is related to the coverage in altitude. Since the data used in the models were taken above a particular altitude, no actual measurements of density below these altitudes could be used. The model predictions at lower altitudes are thus extrapolations of the model under the assumption of hydrostatic equilibrium. Hydrostatic equilibrium for an individual gas species probably does not occur below ~200 km altitude, except in the case of molecular nitrogen. Thus, these models may be suspect at all altitudes below those at which data were taken, and certainly are unreliable below ~200 km.

Since the launch of these two satellites, a number of additional spacecraft carrying mass spectrometers have been placed into Earth orbit. These satellites, taken in combination, provide a significant improvement in data coverage. For example, measurements of atmospheric density and composition are now available over a wide range of altitudes (~140-600 km), and for a range of F10.7 of  $\sim 75-180 \times 10^{-22} \text{ W/m}^2\text{-Hz}$ . Furthermore, a large data base of upper-atmospheric temperature measurements has been compiled at several ground-based incoherent scatter radar sites; these data provide complementary information in the form of direct measurements of temperature with good local time coverage at particular latitudes over a wide range of solar activity and all seasons.

The mass spectrometer-incoherent scatter (MSIS) model (Hedin et al., 1977a, b) makes use of measurements of upper-atmospheric composition from mass spectrometers on five satellites (AE-B, OGO-6, San Marco 3, Aeros A and AE-C)



Table 1. Mean values and standard deviations of  $f$  for ESR0-4/MSIS and OGO-6/MSIS models.

Altitude (km)	N <sub>2</sub>			O			Ar			He		
	$\bar{f}$	Std. dev.	Factor of 2,%*	$\bar{f}$	Std. dev.	Factor of 2,%	$\bar{f}$	Std. dev.	Factor of 2,%	$\bar{f}$	Std. dev.	Factor of 2,%
ESR0-4/MSIS												
200	0.82	0.09	0.0	0.75	0.10	0.0	0.76	0.21	1.3	0.87	0.25	1.4
250	0.82	0.12	0.0	0.74	0.11	0.0	0.76	0.31	8.1	0.86	0.24	2.0
300	0.84	0.17	0.4	0.75	0.11	0.0	0.81	0.41	20.1	0.86	0.24	1.3
350	0.88	0.25	2.6	0.76	0.14	0.0	0.84	0.49	29.5	0.86	0.23	1.2
OGO-6/MSIS												
400	1.16	0.28	1.3	0.99	0.18	0.0	--	--	--	1.11	0.34	3.5
450	1.27	0.35	3.3	1.04	0.20	0.0	--	--	--	1.12	0.34	3.9
500	1.36	0.41	6.9	1.09	0.23	0.1	--	--	--	1.14	0.35	3.7
550	1.43	0.44	10.8	1.15	0.27	0.3	--	--	--	1.15	0.35	3.5

\*The last column for each gas gives the values of  $f$  that deviate by a factor of 2 or more from the respective value of  $\bar{f}$ .

and neutral temperatures inferred from incoherent scatter measurements at four ground stations (Arecibo, Puerto Rico; Jicamarca, Peru; Millstone Hill, Mass., and St. Santin, France). The overall data set covers the time period from the end of 1965 to mid-1975 (effectively a complete 11-year solar cycle), values of  $F_{10.7}$  of 75-180, values of  $A_p$  up to ~100, and is most applicable to the altitude range 200-600 km.

The model formulation is virtually identical to that of the OGO-6 and ESR0-4 model. The MSIS model is more versatile than its predecessors in that the N<sub>2</sub> density, temperature and its vertical gradient are allowed to vary at the 120 km level, the model may be applied to periods of moderate geomagnetic disturbance, a time-independent north-south asymmetry is allowed in the densities, and some allowance for noticeable departures from hydrostatic equilibrium in the diurnal and geomagnetic activity components of the density below 200 km is made.

The various atmospheric species can be crudely ranked as to the absolute error in the model representation of the mean density profile above 150 km: N<sub>2</sub> (~15%), O (~20%), He (~20%), Ar (~25%), H (~50%), and O<sub>2</sub> (~50%). Fortunately, the atmospheric gas above 150 km is composed primarily of those species whose errors are the lowest. At 120 km, the total density is 30% below that of the Jacchia (1971) model, primarily because of the lower N<sub>2</sub> density, which is closer to that measured by rocket-borne mass spectrometers than is that in the Jacchia (1971) model. At altitudes above 205 km, the MSIS model average density is about 15% above that of the Jacchia (1971) model, well within errors of absolute calibration and analysis for both mass spectrometers and orbital drag studies.

The authors of some density models have compared their models with those developed previously. A particularly comprehensive study was carried out by von Zahn and Fricke (1978), who used a statistical approach to compare at various altitudes the number densities of the species N<sub>2</sub>, O, Ar and He as

predicted by the ESR0-4, OGO-6 and MSIS models. All times of the day and year, latitudes and geomagnetic activity index ( $A_p \leq 13$ ) were sampled; comparisons between the ESR0-4 and MSIS models were made at altitudes  $< 350$  km with F10.7 in the range 70-120 (values appropriate for the ESR0-4 data base), while the OGO-6 and MSIS models were compared at higher altitudes and with F10.7 in the range 120-170. Table 1 presents the results of the comparisons as reported by von Zahn and Fricke. The ratios of predicted densities ESR0-4/MSIS and OGO-6/MSIS are shown for the four gases as a function of altitude, along with the standard deviation and the percent of the ratios which deviate by a factor of two or more from the mean ratio. The ESR0-4 model clearly predicts species densities lower (by 12-26%) than MSIS, while the OGO-6 model generally predicts densities higher than MSIS. The ESR0-4 and MSIS models agree better with respect to the standard deviation of the predicted densities. The systematic increase in the OGO-6/MSIS ratio with altitude was attributed to lower temperatures in the MSIS model. The distributions of the ratios were found to deviate substantially from a normal distribution. The percent of ratios differing by more than a factor of two is a measure of this effect; it shows relatively good agreement for O, the major constituent under most conditions treated, but substantial disagreement at the higher altitudes in each case for  $N_2$  and Ar. Further comparison of seasonal and diurnal variations predicted by the models was given, with the ESR0-4 and MSIS models usually in better agreement than OGO-6 and MSIS. Jacchia (1978) has compared the Jacchia (1977) model with these three models and found generally good agreement. The Jacchia (1977) model disagrees with the other models to no larger extent than they disagree among themselves. Jacchia (1978) does point out that the MSIS model, with its treatment of the semi-diurnal variation, is more accurate with regard to the daily variation of the atmospheric constituents below  $\sim 200$  km altitude. On the other hand, the Jacchia (1977) model is probably more accurate in its treatment of the latitudinal dependence of density variations associated with geomagnetic activity.

## 2. COMPARISON OF MODEL PREDICTIONS WITH OBSERVATIONAL DATA

A large number of investigators have compared data from a number of sources with density and composition predictions made by empirical models of the upper atmosphere. The observational data come from both rocket and satellite experiments and consist of total atmospheric density inferred from accelerometer and satellite drag data as well as direct measurements of atmospheric composition from mass spectrometers. A great many comparisons of this type have been made using relatively small amounts of data from various experiments. Although space does not permit a complete review of such studies, a representative sample of recent work will be described. Several detailed statistical studies using large data bases have also been carried out. These investigations will be described in some detail, emphasis being placed on indications of one model's superiority over others and on areas in which models can demonstrably be improved.

A major problem which arises in these comparisons is the choice of a meaningful quantity to calculate. Most investigations have dealt with the ratio  $f$  of the measured density to the model density. The average value



of this ratio, then, is a measure of mean deviation of the model with respect to "actual" densities. This quantity, however, may only represent systematic errors in the measurements; from a user's point of view, such a systematic error, if it lies in the model, may be inconsequential because it may be corrected for, once known. On the other hand, the deviations of  $f$  from its mean value are really more relevant, for they indicate the ability of the model to predict variations in upper-atmospheric structure caused by various geophysical phenomena.

Although the emphasis here is on the ability of models to predict density and composition as measured by instruments other than those whose data were used to build the model, the variability of  $f$  is illustrated by the discussion given by Hedin et al. (1974) regarding the characteristic of the data on which the OGO-6 model is based. Even here, when  $\bar{f} \approx 1.0$ , the value of  $f$  for individual data points for the  $N_2$  density ranges from  $\sim 0.5$  to  $\sim 2.0$ .

With these points in mind, we present here a short review of comparisons of observational data with empirical models. A discussion of studies of total density will be followed by one dealing with composition. Marcos et al. (1977) described total density data at 140 km altitude taken using the MESA accelerometer on the AE-C spacecraft during the period April-December 1974. They compared their data with the CIRA-1972 model and report an average value of  $f$  of 1.07, with a standard deviation of 0.11. Instantaneous values of  $f$  lie in the range  $\sim 0.7$  to  $\sim 1.5$ . The largest values of  $f$  are correlated with large values of  $K_p$ . A latitude dependence of  $\bar{f}$  is evident in their data, with  $\bar{f}$  increasing from  $\sim 1.0$  equatorward of  $45^\circ$ , to 1.17 in the latitude range  $60$ - $68^\circ$ . (The inclination of the orbit is  $68^\circ$ ). An even clearer systematic variation of  $\bar{f}$  with invariant latitude  $\Lambda$  is noted, with  $\bar{f} = 0.98$  at  $0^\circ < \Lambda < 40^\circ$ ,  $\bar{f} = 1.07$  at  $40^\circ < \Lambda < 50^\circ$ ,  $\bar{f} = 1.10$  at  $50^\circ < \Lambda < 60^\circ$ ,  $\bar{f} = 1.14$  at  $60^\circ < \Lambda < 70^\circ$ , and  $\bar{f} = 1.25$  at  $70^\circ < \Lambda < 80^\circ$ . Similar conclusions were given by Rugge (1974) on the basis of accelerometer data from the spacecraft 1970-48A and 1970-61A. These data suggest that a correction to the CIRA-1972 model which involves geomagnetic coordinates may be valuable; this suggestion is substantiated by other studies to be discussed below.

Accelerometer data taken in the latitude range  $\pm 30^\circ$  by the CASTOR satellite during 1975 was reported by Barlier et al. (1977). For data taken below 400 km at local times between 1200 and 2000, comparison with the CIRA-1972 model yielded an average value of  $f$  of 0.82, with variations in the range 0.69 to 0.89. Geographical (latitude, longitude) variations of  $f$  were noted, and some geomagnetic control was noted. (In particular, a region of enhanced  $f$  appears to be very near the location of the South Atlantic Magnetic Anomaly.)

Champion et al. (1974) reported total densities derived from studies of atmospheric drag on the orbits of Cannon Ball 2 (1971-67C) and Musket Ball (1971-67D) in the altitude range 121-158 km. Comparisons of the data with CIRA-1972 yielded mean values of  $f$  extraordinarily near 1.0: 0.996 and 0.999, with standard deviations of .116 and .077. The fluctuations of  $f$  about the mean values appear to have some (but not high) correlation with the values of  $K_p$  and  $F_{10.7}$ . For the Cannon Ball 2 data, a variation of  $f$  with latitude was noted, with relatively large values of  $f$  occurring near the poles and small values near the equator. It is also of interest to note that, although Champion et al. (1974) found  $\bar{f} \sim 1.0$ , Marcos et al. (1977) and Rugge (1974) found  $\bar{f} > 1$ , while Barlier et al. (1977) found  $\bar{f} < 1$ .

More detailed comparisons of the local time dependence of measured total density with models have been carried out using data from San Marco 3 (Brog-

lio et al., 1976) and AE-E (Sharp et al., 1978). Since both of these spacecraft were in low-inclination orbits, the latitude dependence of density could be considered relatively unimportant in the interpretation of the analyses. Broglio et al. (1976) reported that density data taken using a drag balance experiment in the altitude range 200-350 km indicated that the CIRA-1972 model does not represent variations with local time very well, especially at the lower altitudes. The parameter R, which gives the amplitude of the diurnal variation, was found to be altitude-dependent, increasing from a value  $\sim 0.15$  at 200 km, to a relatively constant value of  $\sim 0.23$  above  $\sim 250$  km altitude. This compares with the constant value 0.3 used in CIRA-1972. Furthermore, the data indicate a significant contribution to the diurnal variation from 8 and 6 hour harmonics, and the presence of two diurnal maxima at 205 km was noted.

Sharp et al. (1978) compared total density data taken by the cold cathode ion gauge on the equatorial satellite AE-E in the altitude range 140-300 km with the predictions of the CIRA-1972 and MSIS models. The data show a transition from a predominantly semidiurnal (12 hour) density variation below  $\sim 180$  km to a diurnal (24 hour) variation at higher altitudes. Although the CIRA-1972 model shows no evidence of this behavior, the MSIS model predicts this variation quite well.

Barlier et al. (1976), in a description of a method that might be used to improve the accuracy of empirical models of total density, presented a comparison of satellite drag data from 89 satellites at altitudes in the range 180-500 km with predictions of the CIRA-1972 model. Individual values of  $f$  ranged from  $\sim 0.48$  to  $\sim 2.19$  and had a distribution close to log-normal about a mean value near 1.0. Systematic variations of  $f$  with parameters such as latitude, local time,  $F_{10.7}$ , altitude and  $A_p$  were identified.

Hickman et al. (1978) compared  $\sim 32000$  in-situ measurements of total density in the altitude range 135-300 km made by cold cathode ion gauges on board the AE-C, -D and -E spacecraft with the predictions made by several empirical models. Evaluation of  $f$ , the rms error  $\sqrt{f^2}$ , and the standard deviation  $\sqrt{(f - \bar{f})^2}$  was carried out; the first and third of these quantities are plotted here in Figure 1 for the CIRA-1972, MSIS and Jacchia (1977) models. The overall averages of  $f$  for these models are 1.036, 0.966 and 1.029, respectively; systematic variations of  $f$  with altitude are also evident. The values of  $\bar{f}$  for CIRA-1972 and J-77 are very similar; the additional complexity of Jacchia (1977) has not led to improved predictions of total density. The rms deviations from the mean are 0.191, 0.179 and 0.196 for the three models and show weak altitude dependences.

The preceding discussion has been limited to total density predictions. Mass spectrometer data which bear on the composition of the upper atmosphere have been available in quantity since the launch of OGO-6 in 1969. Since then, substantial amounts of data have been obtained by instruments on San Marco-3, ESRO-4, and the AE satellites. These data indicate that earlier static models with constant boundary conditions near the turbopause altitude cannot accurately represent the seasonal, diurnal, and geomagnetic activity-related variations of all the atmospheric constituents, especially in the lower thermosphere ( $\sim 100$ - $\sim 200$  km). The requirement that predictive models represent the composition variations (which are largely hidden when only total density is measured) has led to increased complexity of newer models, as discussed earlier. Here we shall assess the predictive capability of some of these models, as evidenced by comparison with mass spectrometer data from several spacecraft. As in the case of total density, most comparisons



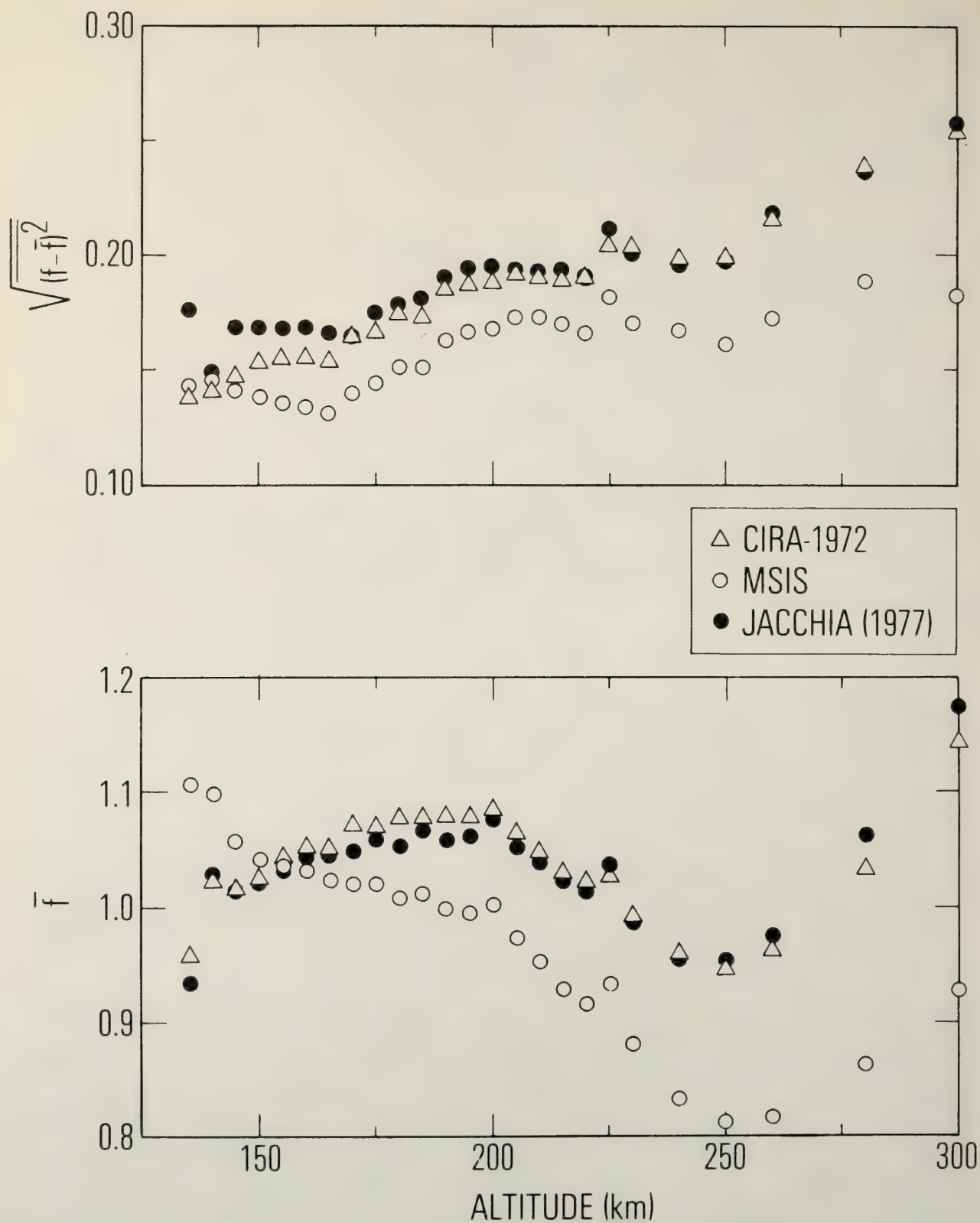


Figure 1. Values of  $\bar{f}$  and  $\sqrt{(f-\bar{f})^2}$  for the CIRA-1972, MSIS, and Jacchia (1977) models.

have been made with relatively small data sets; such studies generally cannot isolate systematic variations. We shall concentrate here on more large-scale investigations, whose results might lead to improved predictive capability.

Kasprzak and Newton (1976) compared composition measurements of the San Marco 3 magnetic mass spectrometer below 280 km with predictions of the OGO-6 model. Since San Marco 3 was in a low inclination orbit, local time variations could be clearly discerned. In general, the measured total oxygen ( $O + 2O_2$ ) density and OGO-6 model predictions were found to agree to within  $\pm 30\%$ , while agreement for  $N_2$  was within  $\pm 40\%$ . For both of these constituents, systematic local time variations in  $f$  of  $\sim \pm 15\%$  were noted. Little systematic altitude variations were seen. Agreement between measured and model He densities was considerably worse:  $f$  ranged from  $\sim 0.3$  to 5, with systematic local time dependence (low values of  $f$  during the day and large values at night) and altitude dependence ( $f$  decreasing with altitude). The points of disagreement between measured local time variations of the composition and the OGO-6 model were attributed to the fact that the OGO-6 model involves an extrapolation of data taken at  $\sim 450$  km using diffusive equilibrium assumptions. The data indicate a strong deviation from diffusive equilibrium, especially for He, below  $\sim 200$  km altitude. It was also noted that the CIRA-1972 model, which has fixed lower boundary conditions for both temperature and composition, accurately predicts neither the observed diurnal variation amplitude nor the observed phase of the He density variation.

Fricke et al. (1976) compared 8000 composition measurements from the gas analyzer on the ESR0-4 satellite in the altitude range 250-300 km with the predictions of the CIRA-1972 model. They found that, for total density,  $\bar{f} < 1$  at these altitudes, with the value of  $\bar{f}$  increasing from  $\sim 0.8$  at 300 km to  $\sim 0.95$  at 250 km. For atomic oxygen,  $\bar{f} \approx 0.66$  at 275 km, whereas  $\bar{f}$  for  $N_2$  is  $\sim 1.99$ ; thus the overestimate of  $O$  in the model is partially offset by the underestimate of  $N_2$ , leading to a total density prediction only  $\sim 10\%$  too high at 275 km altitude. Whereas the variations of the measured densities about their mean values for  $O$  and  $N_2$  are nearly normal, the variations for He and Ar are quite skewed, with long tails toward higher values. On average, helium densities are seriously overestimated by CIRA-1972, whereas argon densities are underestimated. Seasonal variations in the  $O/N_2$  ratio clearly demonstrate the effects of thermospheric circulation on the distributions of these constituents. The CIRA-1972 model, although reasonably accurate for predictions of total density, does not distribute observed total density variations correctly among the various species.

Von Zahn and Fricke (1978) carried out a very detailed comparison of composition data taken by the ESR0-4 gas analyzer at 275 km with the ESR0-4 and MSIS models. Although on average, the ESR0-4 model fits the ESR0-4 data for  $N_2$ ,  $O$ , Ar, and He better than does the MSIS model, the standard deviations of  $f$  for the two models are almost identical (0.25, 0.17, 0.33, and 0.27 for  $N_2$ ,  $O$ , Ar, and He, respectively), indicating that the models represent the composition variations equally well at this altitude, even though no data from ESR0-4 were contained in the MSIS data base. The discrepancies in  $\bar{f}$  for the four constituents (0.87, 0.73, 0.75, and 0.87) for the MSIS model were noted and attributed to uncertainties in the absolute sensitivities of the various mass spectrometers which contributed to the MSIS data base. The major contribution of von Zahn and Fricke (1976) lies in their investigation of the density residuals that arise in comparing the ESR0-4 measurements with the ESR0-4 model. In addition to transient phenomena, such as gravity



waves, which affect the density in a way which is effectively unpredictable, there appear to be systematic features of thermospheric structure that are not contained in present models. Von Zahn and Fricke demonstrated variations in the four atmospheric constituents that are under geomagnetic control. Figure 2 shows their analysis of  $\log f$  as a function of longitude and geographic latitude for the four constituents. A clear concentration of values of  $f > 1$  near the geomagnetic poles is evident for the heavier species  $N_2$  and Ar, whereas  $f < 1$  generally near the geomagnetic poles for O and He. A suggested improvement in the model, then, is the addition of a sinusoidal variation in longitude for the individual gas densities whose amplitude and phase depend on latitude. A similar improvement could be achieved by using both geographic and geomagnetic coordinate systems in future models.

In addition to in-situ measurements of thermospheric composition by earth-orbiting spacecraft, incoherent scatter radar provides a ground-based method for measuring a number of quantities of relevance here. Alcayde et al. (1974) compared long-term variations of daytime observations at  $\sim 45^\circ N$  latitude of the atomic oxygen concentration in the 200-400 km altitude range and the ratio of atomic to molecular species at  $\sim 200$  km with predictions made by the CIRA-1972 model. Two years of data for the atomic oxygen concentration at 200 km indicate that the CIRA-1972 model is, on average, quite good, but that variations in  $f$  for atomic oxygen ranging from  $\sim 0.3$  to  $> 3.0$  occur. A semi-annual variation in  $f$  is evident, with maxima near the equinoxes; a decrease in  $f$  with increasing geomagnetic activity is also noted, a behavior consistent with the effects of global-scale circulation changes associated with geomagnetic activity. Data reported by Alcayde et al. also indicate that CIRA-1972 predicts total mass density better than it does the density of individual atmospheric species. Thus, the incoherent scatter measurements lead to the same general conclusions reached on the basis of mass spectrometer measurements: a model with constant boundary conditions near the turbopause and assuming diffusive equilibrium cannot adequately represent variations in atmospheric composition, although it can predict total density with reasonable accuracy.

### 3. CONCLUDING REMARKS

We have reviewed the formulations and characteristics of a number of empirical models of thermospheric density and composition. Today's models are much more accurate (and complex) than were the first standard models of twenty years ago. Models produced prior to 1970 were developed from data bases with significant limitations. Most of the data were from satellites with perigee altitudes above 250 km and were based on drag calculations, yielding only total density. The only advantage these models have over more recent models is their greater simplicity and the many years of experience which some groups have had in using them. More recent models are more suitable except for those situations where only total density is required and some loss of accuracy can be tolerated in exchange for small, fast computer code or where the cost of replacing present software greatly exceeds the benefits achievable by the use of better (but still far from perfect) models.

The best models for general use are probably CIRA-1972 and MSIS. Although CIRA-1972 is more widely used, the MSIS model is particularly appropri-

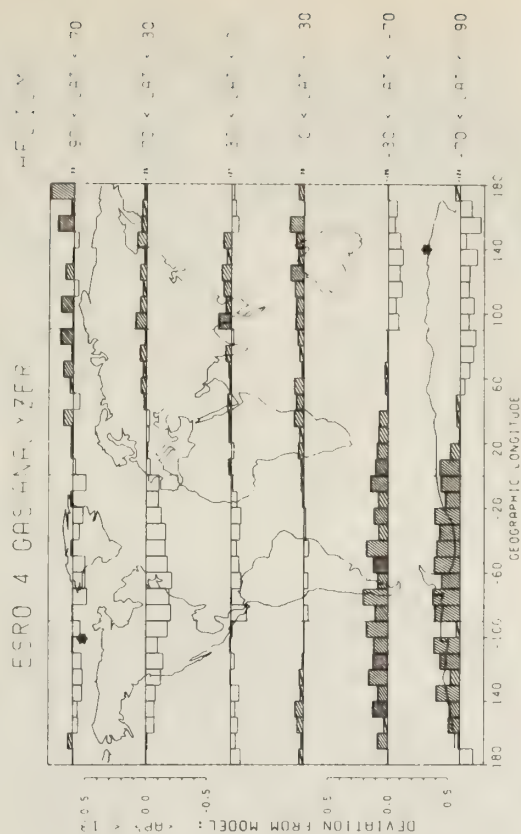
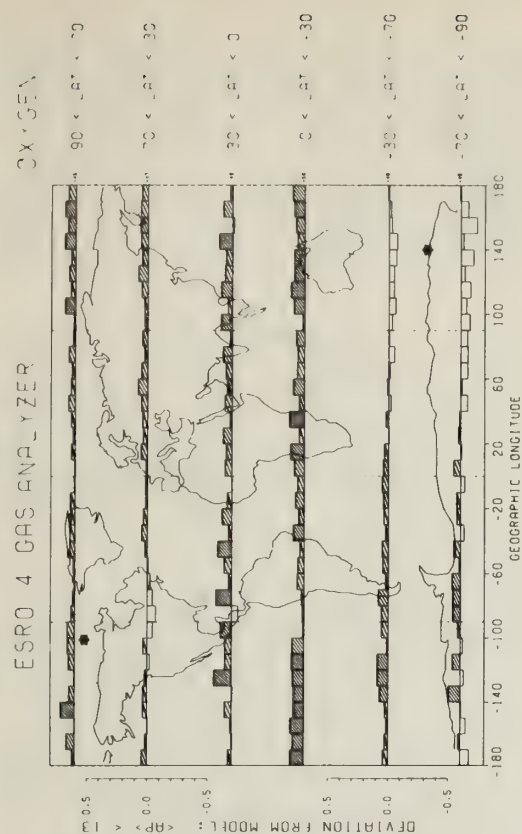


Figure 2. Analysis of  $\log f$  as a function of longitude and geographic latitude for nitrogen, argon, oxygen, and helium. [After von Zahn and Fricke (1978).]



ate when predictions of compositions are required, or when the altitude range of interest extends to below ~200 km. Furthermore, CIRA-1972 requires numerical integration of the diffusive equilibrium equation for each gas; this is a time-consuming procedure. Jacchia's (1977) model, although limited in the amount of low altitude data used in its formulation, does describe the global distribution of the individual species in a manner consistent with recent observations. The major objection to this model is its complexity and resulting inconvenience of use. Jacchia (1978) has recognized this, and a simpler version of the model is forthcoming.

The overall accuracy of the recent models can be summarized as follows. On the average, the models are as good as are measurements of atmospheric density and composition. This is evidenced by the fact that some measurements find  $\bar{f} > 1$ , while others find  $\bar{f} < 1$ . We interpret this situation as being due to systematic instrumental inaccuracies. However, individual measurements of  $f$  show a scatter of about a factor of 2 (above and below) the mean value, with a standard deviation of about 15-20% for the total density and larger values for some of the individual atmospheric constituents.

That a model must be accurate is obvious. Less obvious is how the accuracy is to be characterized for a particular use. A perfect description of the atmosphere is the ideal, but some kinds of inaccuracies are unimportant for specific applications. A satellite system designer who is required to estimate the orbital lifetime until reentry of a satellite designed for a two-year lifetime may be quite satisfied with a model having the correct mean value when averaged over a month's time span even if the errors averaged over a single orbit are quite large. On the other hand, a tracking station operator may attach no importance at all to the long-term mean behavior of the atmosphere (which could be empirically adjusted for in day-to-day operations) if the model could only accurately predict the short-term variations.

Why is it that models do not do a better job of predicting thermospheric density? Part of the answer lies in the selection of parameters to characterize conditions in the thermosphere. For example, the amount of EUV heating is often measured by the 10.7 cm solar flux. The 10.7 cm flux itself cannot significantly heat the upper atmosphere. It has been established, however, that there is a reasonable correlation between the EUV flux and the 10.7 cm flux. This correlation is only approximate and may not be adequate for use in accurate models. However, the 10.7 cm flux is measured regularly and hence it is readily available; better a "fair" index than none at all. A similar situation exists with respect to the energetic particle heating which occurs during geomagnetic storms. This heating is related to disturbances in the Earth's magnetic field, but the relationship is undoubtedly not adequate for high accuracy models. There is a need for new parameters to be made available which more directly characterize phenomena which affect the thermosphere. Although this need has been recognized for over a decade, no significant new parameters have emerged.

As described by some investigators (e.g., Marcos et al., 1977; von Zahn and Fricke, 1978), systematic variations of  $f$  (for both density and composition) with geomagnetic, or invariant, latitude and geographical longitude are evident. Ordering of the data with respect to one of these parameters, which are really equivalent since their use is meant to represent variations induced by geomagnetic phenomena (the geomagnetic dipole axis not being coincident with the earth's axis of rotation), could presumably improve existing models. Furthermore, if accuracy, rather than convenience of use, is the overriding consideration in the development of a model, the assumption of

diffusive equilibrium for the individual constituents should be laid aside, at least below ~200 km altitude. (An initial attempt in this area was made in the MSIS model.) As noted earlier, although the total density is well represented in the models, substantial inaccuracies still exist for the densities of individual constituents there. Such a modification would lead to models of greatly increased complexity, and their use might be warranted only for applications in which very accurate composition predictions are required.

The conventional view of the progress of modeling efforts over a period of time is of a number of evolutionary stages. In the early days there are a number of measurements, but little understanding of the physical processes which are in control. At this stage, one is forced to represent the data by some other general method such as polynomial expansion (in one dimension) or spherical harmonic expansion (for data on a spherical surface). As understanding of the modeled behavior improves, there is a replacement of these equations of general applicability with more specific equations based on the ever-improving understanding of the underlying physics which govern the situation. For empirical thermospheric modeling it is curious to note that the trend seems to be in the opposite direction. Even though theoretical models have led to a significant increase in the understanding of atmospheric dynamics, equations based on a simplified model of some physical process are giving way to descriptions based on spherical harmonic expansions. The overly simplified picture used as the basis for early models has shown itself to be inadequate for use in high accuracy models. Thus we see the retreat to more empirical kinds of formulations. The diffusive equilibrium approach, which cannot accommodate some of the dynamical effects such as are produced by horizontal winds, may be too simple a basis for present-day requirements. An alternate approach, so far not tried, would be to attempt to incorporate a more complex underlying physical model in place of the current diffusive equilibrium equation schemes. Perhaps a rather simple two or three dimensional model, with terms corresponding to horizontal transport and similar effects, could be developed. If such an atmospheric model is to be used for the same purposes as present empirical models, the aim would have to be toward a relatively simple model with good computational efficiency rather than toward inclusion of all known physical processes. Some parameterization of unmodeled effects would still be required.

On the other hand, even if such improvements were made, it is not clear that extremely accurate models are really possible. Gravity waves and other transient disturbances affect the density in a seemingly random fashion. It may be that ultimately we will be able to accurately predict the average state of the atmosphere, but not be able to predict short-term (e.g., on a time scale shorter than a few hours) departures from this average state. In the language of the meteorologist it may be that we can predict the "climate" but not the "weather." It is not clear at present what the ultimate accuracy of models can be. There is still much to be done in developing the theoretical understanding of thermospheric behavior, particularly the interaction of the thermosphere with the mesosphere below and the magnetosphere above.

Although there is no immediate prospect of a revolutionary improvement in the accuracy of thermospheric models, there are, nevertheless, good reasons for believing that prediction of thermospheric behavior may undergo significant improvement in the long run. Unlike tropospheric weather prediction, there does not exist the problem of local geography causing small-scale local differences. Propagation times for disturbances are of the order of hours,



which helps to smooth out perturbations quickly. The coupling between the mesosphere and the thermosphere and the unknown behavior of the atmosphere in the region near 90 km is a problem, as is the lack of routine monitoring of the state of the entire thermosphere by the equivalent of present meteorological satellites which monitor the troposphere and stratosphere. As remote sensing techniques are developed and applied to the mesosphere and lower thermosphere, we may reasonably expect that there will be steady, though probably not spectacular, progress toward high accuracy models.

### Acknowledgments

The authors would like to thank B. K. Ching, C. J. Rice and L. R. Sharp for useful discussions. This work was supported by The Aerospace Corporation under its Research Support Program.

### REFERENCES

- Alcayde, D., P. Bauer, and J. Fontanari (1974): Long-term variations of thermospheric temperature and composition. J. Geophys. Res., 79:629.
- Barlier, F., J. P. Bordet, and J. P. Orfeuill (1976): Methods of analysis of satellite drag density values based upon statistical treatment and factorial analysis. Space Research, 16:161.
- Barlier, F., Y. Boudon, J. F. Falin, R. Futaully, J. P. Vallain, J. J. Walch, A. M. Mainguy, and J. P. Bordet (1977): Preliminary results obtained from the low-g accelerometer CACTUS. Space Research, 17:341.
- Broglio, L., C. Arduni, C. Buongiorno, U. Ponzi, and G. Ravelli (1976): Diurnal density variations measured by the San Marco 3 satellite in the equatorial atmosphere. J. Geophys. Res., 81:1335.
- Champion, K. S. W., D. F. Gillette, and I. M. Hussey (1974): Atmospheric densities from Cannon Ball 2 and Musket Ball satellite observations, Space Research, 14:125.
- CIRA-1961, COSPAR International Reference Atmosphere (1961): North-Holland Publishing Co., Amsterdam.
- CIRA-1965, COSPAR International Reference Atmosphere (1965): North-Holland Publishing Co., Amsterdam.
- CIRA-1972, COSPAR International Reference Atmosphere (1972): Akademie-Verlag, Berlin.
- COESA, U.S. Standard Atmosphere (1962): U.S. Government Printing Office, Washington, D.C.
- COESA, U.S. Standard Atmosphere (1976): U.S. Government Printing Office, Washington, D.C.

- Fricke, K. H., U. Laux, H. Trinks, and U. von Zahn (1976): ESRO-4 gas analyzer results and CIRA-1972 predictions: A statistical comparison. Space Research, 16:265.
- Harris, I., and W. Priester (1962a): Time-dependent structure of the upper atmosphere. J. Atmos. Sci., 19:286.
- Harris, I., and W. Priester (1962b): Theoretical models for the solar-cycle variation of the upper atmosphere. J. Geophys. Res., 67:4585.
- Hedin, A. E., H. G. Mayr, C. A. Reber, N. W. Spencer, and G. R. Carignan (1974): (1974): Empirical model of global thermospheric temperature and composition based on data from the OG0-6 quadrupole mass spectrometer. J. Geophys. Res., 79:215.
- Hedin, A. E., J. E. Salah, J. V. Evans, C. A. Reber, G. P. Newton, N. W. Spencer, D. C. Kayser, D. Alcayde, P. Bauer, L. Cogger, and J. P. McClure (1977a): A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 1. N<sub>2</sub> density and temperature. J. Geophys. Res., 82:2139.
- Hedin, A. E., C. A. Reber, G. P. Newton, N. W. Spencer, H. C. Brinton, H. G. Mayr, and W. E. Potter (1977b): A global thermospheric model based on mass spectrometer and incoherent scatter data MSIS 2. Composition. J. Geophys. Res., 82:2148.
- Hickman, D. R., B. K. Ching, C. J. Rice, L. R. Sharp, and J. M. Straus (1978): A survey of currently important empirical thermospheric models. The Aerospace Corporation.
- Jacchia, L. G. (1960): A variable atmospheric density model from satellite accelerations. J. Geophys. Res., 65:2775.
- Jacchia, L. G. (1965): Static diffusion models of the upper atmosphere with empirical temperature profiles. Smithsonian Contrib. Astrophys., 8:215.
- Jacchia, L. G. (1970): New static models of the thermosphere and exosphere with empirical temperature profiles. Smithsonian Astrophys. Obs. Spec. Rep. No. 313, Cambridge, Mass.
- Jacchia, L. G. (1971): Revised static models of the thermosphere and exosphere with empirical temperature profiles. Smithsonian Astrophys. Obs. Spec. Rep. No. 332, Cambridge, Mass.
- Jacchia, L. G. (1977): Thermospheric temperature, density, and composition: New models. Smithsonian Astrophys. Obs. Spec. Rep. No. 375, Cambridge, Mass.
- Jacchia, L. G. (1978): CIRA-1972, recent atmospheric models and improvements in progress. Paper presented at 21st meeting of COSPAR.



- Kasprzak, W. T., and G. P. Newton (1976): Comparison of the San Marco 3 Nace neutral composition data with the extrapolated OGO-6 model. J. Geophys. Res., 81:1404.
- Marcos, F. A., K. S. W. Champion, W. E. Potter, and D. C. Kayser (1977): Density and composition of the neutral atmosphere at 140 km from atmospheric Explorer-C satellite data. Space Research, 17:321.
- Minzner, R. A. (1977): The 1976 standard atmosphere and its relationship to earlier standards. Revs. of Geophys. and Space Phys., 15:375.
- Minzner, R. A., K. S. W. Champion, and H. L. Pond (1959): The ARDC model atmosphere, 1959. Air Force Surveys in Geophysics No. 115, ARCRC-TR-59-267.
- Rugge, H. R. (1974): Comparison of densities deduced from radar determined orbital decay and accelerometer measurements. Space Research, 14:131.
- Sharp, L. R., D. R. Hickman, C. J. Rice, and J. M. Straus (1978): The altitude dependence of the local time variation of thermospheric density. Geophys. Res. Lett., 5:261.
- von Zahn, U. (1970): Neutral air density and composition at 150 kilometers. J. Geophys. Res., 75:5517.
- von Zahn, U., and K. H. Fricke (1978): Empirical models of global thermospheric composition and temperature during geomagnetically quiet times compared with ESR0-4 gas analyzer data. Rev. Geophys. and Space Phys., 16:169.

SOLAR-INDUCED VARIATIONS OF ENVIRONMENT AFFECTING  
MANNED SPACE FLIGHT AND SPACECRAFT OPERATIONS

William G. Johnson, Thomas A. Parnell,  
and William W. Vaughan  
Space Sciences Laboratory  
NASA George C. Marshall Space Flight Center  
Huntsville, Alabama 35812, USA

Emphasis in the paper is upon the induced variation of the environment in the areas of spacecraft charging, radiation effects, and orbital lifetime together with an evaluation of current prediction value and future needs.

Spacecraft operating at geosynchronous orbital altitudes have experienced operational anomalies due to electrical discharges that occurred when dielectric surfaces of the vehicle were charged by interaction with the ambient plasma to levels above breakdown voltages. Correlations between the time of occurrence of the anomalies and solar activity appear to exist. As the size of spacecraft operating at geosynchronous altitudes increases, the present practice used to avoid anomalies may not be economically feasible. An ability to accurately predict the solar-induced variability of the plasma environment at spacecraft orbiting altitudes could permit refinement of design criteria and, possibly, development of operational techniques to control deleterious results of anomalies.

The effects of the energetic particle environment on spacecraft operations generally increase with the orbit altitude and inclination, as do solar-induced variations in the radiation environment. The influence of the radiation on spacecraft materials, systems, and manned operations is summarized. Areas, such as solar flare models, where improvement in predictions would reduce the impact on operations are discussed.

Spacecraft lifetimes are a direct function of the short- and long-term variations in the orbital altitude atmospheric density environment. These variations are due to the heating of the Earth's upper atmosphere by solar radiation. The effect of density on the orbital lifetime increases with decreasing orbital altitude. The importance of these variations is found in the requirement for orbital performance capabilities which will insure design lifetime, definition of orbital dynamics for nonspherical spacecraft, and the assessment of lifetime potential for spacecraft in orbit. Estimates of short- and long-term solar activity levels are critical input to these calculations.



## INTRODUCTION

This paper is divided into sections according to the principal topics. The first section provides a description of the major problems and needs for solar predictions associated with spacecraft operational anomalies related to electrical discharges when dielectric surfaces of the vehicle are charged by interaction with the ambient plasma to levels above the breakdown voltages. The second section treats the effects of cosmic radiation on spacecraft materials, systems, and operations. Solar influences and the need for improved predictions of solar activity will be presented and areas noted where satellite and spacecraft operations can be improved. The third section discusses the need for improved solar activity predictions relative to the estimation of spacecraft orbital lifetimes. The last section summarizes the need for more accurate solar activity predictions to serve the requirements of the preceding three problem areas associated with spacecraft design and operations.

## SOLAR PREDICTIONS AND SPACECRAFT CHARGING

A discussion of possible relationships between solar states and the charging of spacecraft in high-altitude orbits would be speculative. This is not a totally surprising condition. It is similar to that which now exists in the area of solar-weather relationships. In that field, it currently appears reasonable to assume that there are causative, connective chains relating the state of the Sun to the state of terrestrial weather. A large number of correlations have been suggested, and at least portions of a number of causative chains have been proposed. Yet, currently no single such chain has been validated in its entirety. Further, as more data have become available, and as our understanding of the system has increased, many of the correlations which initially seemed significant have proved to be not well founded. In the case of spacecraft charging, DeForest (1977) has provided an excellent evaluation of our current state of knowledge: "We are still finding new plasma phenomena at geosynchronous orbit. We understand the overall patterns fairly well and are making progress on understanding such things as waves. But one must remember that this is a very complex environment." He observes in connection with changes in the environment that result in observed anomalies in operations of the spacecraft: "A convenient comparison is to say that substorms are like the earthly thunderstorms that we can predict and understand reasonably well. Many of these unusual events are like tornadoes. We know they are associated with larger events and they are potentially dangerous."

At altitudes above 120 km, the electrical characteristics of the atmosphere are, perhaps, the characteristics of greatest interest. From the earliest days of space research it was recognized that measurements of the ambient environment from an orbiting spacecraft were contaminated by the effects of motion of the spacecraft through the atmosphere and by the effects of the electrical state of the spacecraft on the environment. For typical treatments see, for example, Davis and Harris (1961), Singer and Walker (1962), Parker and Whipple (1970), Parker (1976), and Whipple *et al.* (1974). Hence, the state of charge of an orbiting spacecraft has been a long-standing area of interest. However, as used in this discussion, "spacecraft charging" shall be defined to be essentially the inverse of the problem alluded to previously. It is the effect of the environment on the electrical behavior of a spacecraft.

Fredricks and Scarf (1973), reporting the results of an investigation of operational anomalies occurring in spacecraft in geosynchronous orbit, identified a class of anomalies which appeared to be clearly associated with the interaction of the spacecraft with the space plasma. McPherson, Cauffman and Schober (1975) showed the time of occurrence of the events to be strongly correlated with local time in orbit, predominantly occurring in the midnight to dawn quadrant. An analysis of ATS-6 spacecraft charging events by Reasoner et al. (1976), revealed a similar distribution in time. These observed correlations suggested as a physical explanation of the observed anomalies the coupling of the electromagnetic radiation, generated when electrical discharges occurred between differentially charged external surfaces of the spacecraft, into the sensitive bistable logic circuits. The differential charging occurred when the dissimilar exterior surfaces were exposed to "clouds of hot plasma" in the region of the magnetosphere tail.

Inouye (1977), reporting an analysis of a more extensive set of anomalies of the types reported by Fredricks and Scarf, failed to find the strong correlation with local time in orbit. He found, at best, a weak correlation of certain subsets of anomalies with magnetic activity indexes. The strongest correlation was found to be with day of the week, with a "significant" increase in the rate of occurrences taking place during the week end. Inouye suggests two possible causes for the observed effect, one associated with the change in usage rate of the spacecraft on the week end and the other with influences of the Canadian Power Network on the magnetosphere.

Robbins and Short (1977) have analyzed anomalies occurring on the Skynet 2B spacecraft during 1975-1976. They conclude that the anomalies were due to electromagnetic rf interference generated in the vicinity of the spacecraft. They found no correlation between the rate of occurrence of the anomalies and local time in orbit and only a weak correlation with geomagnetic activity. They did observe a relatively significant increase in the rate of occurrence during the eclipse seasons which they suggest "may be associated with photoemissive effects and changes in spacecraft Sun angle."

An analysis of the RCA Satcom 3-Axis Spacecraft operational experience by Napoli and Seliga (1977) failed to reveal "any anomalous behavior that can be attributed to the effects of spacecraft charging." They identify the precautions that were taken in the design and development of the spacecraft which they believe reduced the potential for uncontrolled charge buildup on the exterior surfaces and increased the immunity of internal circuits to EMI should external discharges have occurred.

Baughner (1978), on the basis of preliminary analysis of approximately 9 months of ISEE operational data, finds no charging events of the magnitude observed on ATS-6. The ISEE spacecraft were designed with the ATS-6 experience in hand and with an objective of reducing uncontrolled influences of the carrier vehicle on the environment. Hence, control of charge accumulation on the spacecraft surface was a primary design consideration.

The precise criteria that a spacecraft is to be designed to meet are determined invariably by compromise. Although frequently not definitely enunciated, cost is always a consideration of primary concern. Recent experiences appear clearly to indicate that much of the information needed to



design spacecraft that are immune to upset because of charge accumulation is now available. In the case of smaller spacecraft, it appears that many of the design practices that reduce charge buildup (e.g., interconnection of thermal protection blanket reflective layers, conductive coating on solar-cell covers, conductive thermal control paints, etc.) and of those that increase immunity of internal electronic circuits to surface discharge-induced EMI (e.g., separation and shielding of signal cables, layouts to reduce signal cable lengths, single-point grounding, etc.) can be economically applied.

NASA is currently contemplating spacecraft to be operated at geosynchronous orbital altitudes in which solar cell array areas will be measured in thousands of square meters, in which thermal control systems will be of similar sizes, in which command and control functions are to be executed from widely distributed points and signal cable runs may be measured in hundreds of meters, and in which voltages and power densities will be many orders of magnitude greater than on present spacecraft. It is not yet obvious that designs which would be justified for smaller sized vehicles can be economically applied in toto to spacecraft of the sizes envisioned. Hence, it seems probable that the protection of these future spacecraft will involve combinations of the protective designs that can be economically incorporated and of operational procedures to reduce the consequences of surface charge accumulations.

During the past decade, as more data have become available on the occurrence of spacecraft charging events and on the behavior of the ambient environment, the ability to hypothesize satisfying causative mechanisms appears to have decreased. The relatively simple model advanced to explain the early observations of charging events reported by Fredricks and Scarf (1973) does not sufficiently explain the more extensive observations reported by Inouye (1977) and by Robbins and Short (1977). Models of the environment and its behavior have increased in complexity. Certainly, we are not now at a point at which we can predict, on the basis of observation of the state or behavior of the Sun, the behavior of the magnetosphere and plasmasphere that would lead to a spacecraft charging event. Nevertheless, it does not appear irrational to assume the existence of causative chains relating solar behavior to the electrical behavior of the atmosphere at geosynchronous orbital altitudes. The benefits that would accrue to the operation of complex spacecraft in that region if our understanding of the behavior of the environment were sufficient to permit a predictive capability appear evident.

## SOLAR PREDICTIONS AND THE RADIATION ENVIRONMENT

The ubiquitous radiation in space, now reasonably well known in particle content, energy spectrum, and spatial extent, continues to be a principal consideration in both manned and unmanned spacecraft. Even though the radiation environment external to and within spacecraft has received extensive analytical treatment and measurements (Burrell et al., 1968; Warman, 1971; Watts and Wright, 1976) and radiation effects on materials components and man have been studied in considerable detail (Hill, 1978), the radiation problem still must be considered on an ad hoc basis for each mission. The reason for the continued necessity for an evaluation of radiation effects is threefold. First, new and unanticipated effects of the radiation frequently arise (May and Woods,

1978; Fishman and Austin, 1976). Second, the radiation varies rapidly with orbital parameters (not only in intensity, but in particle types and prominent effects). Third, temporal variation in the radiation environment, caused by solar activity, increases the complexity of predicted radiation levels for many space missions. All three of these factors reduce the effectiveness of a cookbook approach to predicting radiation effects. This section presents a brief survey of the main constituents of the radiation environment and some of the more obvious radiation effects, with emphasis on the constituents that vary with solar activity.

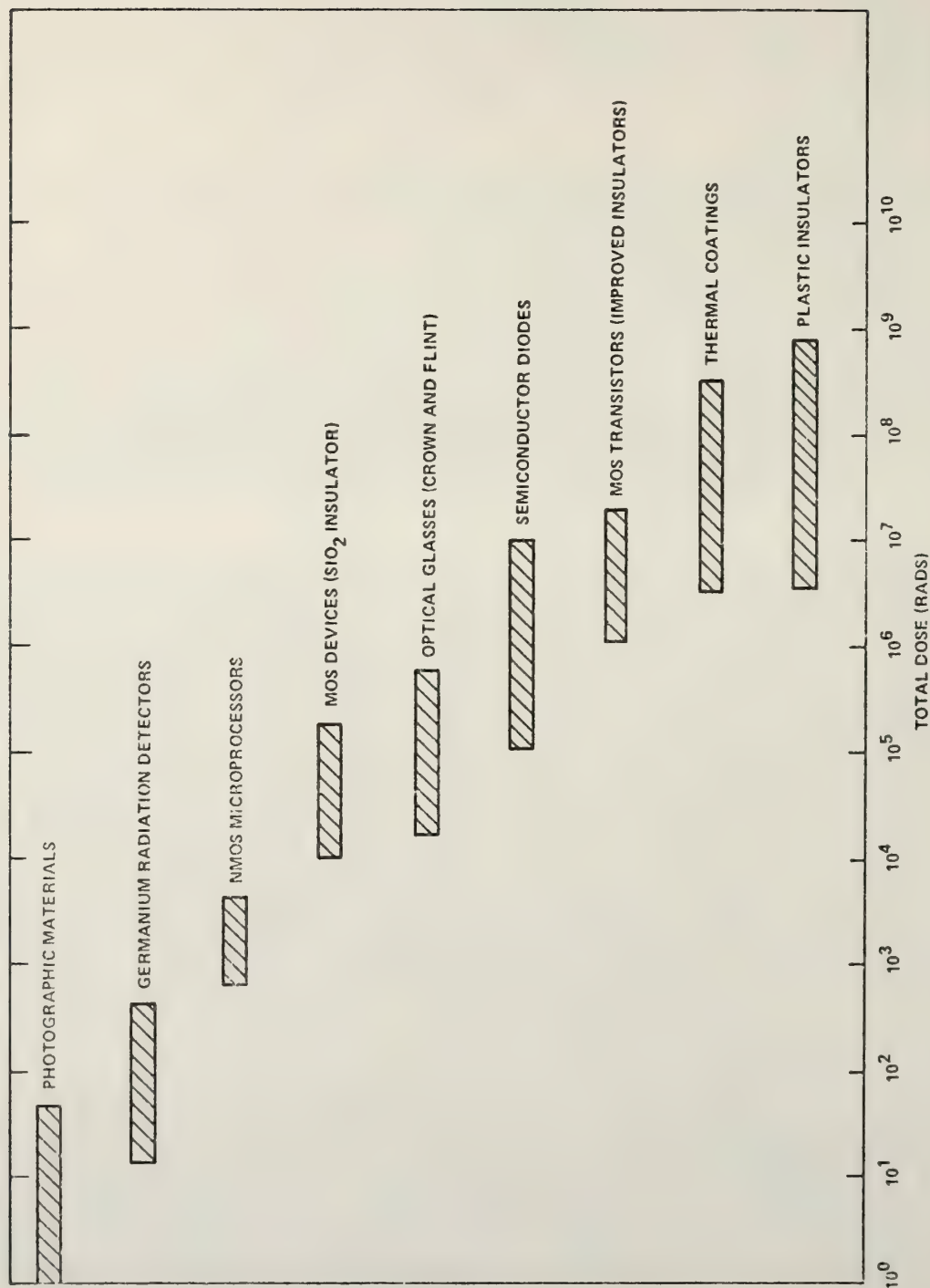
The effects of radiation on materials, components, and man vary widely and cover a large range in radiation dose and particle flux. The deleterious effect of the radiation may depend on either total dose, dose rate, particle identity and energy, or a combination of these. The effect may vary from background noise in a sensor at a low particle flux to the change in characteristics of thermal control surfaces at  $\sim 10^9$  rads total surface dose. In Figure 1 (adopted from Hill, 1978) some accumulated data are presented showing the approximate ranges of radiation dose that affect a number of common materials and devices. Such a presentation does not carry information on variations due to the factors mentioned previously but indicates the dose range for which further analysis must be carried out in mission design and planning.

An example of sensitivity to a particular radiation constituent are the various effects of the heavy nuclei which are present in the galactic cosmic ray flux and in solar flares. Although these particles carry only a small fraction of the total dose (in rads) at all altitudes up to geosynchronous, some effects are quite pronounced. These particles have recently been shown to produce a significant rate of transient errors in digital logic (Binder et al., 1975; May and Woods, 1978), and this effect will influence further miniaturization of large-scale integrated circuitry for use in space. These particles also produce significant noise in certain types of sensors (Fishman and Austin, 1976), and their influence on man and other biological systems has not yet been fully assessed (Benton et al., 1977).

Table 1 shows radiation limits previously established for manned operations in space. The dose limits are established in rems (radiation equivalent, man) in which greater significance (quality factor) is given the heavy particles. Thus, a dose expressed in rads will be somewhat higher when converted to rems for space radiation. These limits are set sufficiently high that they have not imposed significant restrictions on previous manned space operations, and this will probably continue to be the case for missions in low-altitude, low-inclination orbits and stay times of a few months. However, the trapped belt radiation varies rapidly with altitude and inclination of the orbit, and planning for many future missions will be concerned with operational limitations due to radiation. Also, at inclinations greater than  $\sim 50$  degrees at low altitudes and for all inclinations at altitudes of  $>5$  Earth radii, the impact of solar activity, especially solar flares, becomes of particular concern.

Considering now the radiation dose that can be anticipated in a typical spacecraft situation, Table 2 gives some representative examples from calculations. For a large space structure the dose calculations for  $5 \text{ gm/cm}^2$  of





NASA/NSFC/ES-61

FIGURE 1. Approximate range for sensitivity of example components to the ionization effects of radiation.

TABLE 1: Suggested exposure limits (rem)\*

<u>Constraint</u>	<u>Bone Marrow</u>	<u>Skin</u>	<u>Ocular Lens</u>	<u>Testes</u>
Avg. Daily Rate	0.2	0.6	0.3	0.1
30 Days	25	75	37	13
Yearly	75	225	112	38
Career	400	1200	600	200

\* NRC Publication 1487, "Radiation Dose Levels for Apollo Crew Members,"  
File Memo FA, 2-10-67, 1967

TABLE 2: Predicted radiation dose for one year exposure (rads)\*\*

<u>Orbit</u>	<u>0.1 gm/cm<sup>2</sup> Shielding Trapped + Cosmic Ray</u>	<u>5.0 gm/cm<sup>2</sup> Shielding Trapped + Cosmic Ray</u>	<u>Solar Flares<sup>†</sup></u>	<u>Total</u>
400 km/30°	400	32	--	32
400 km/90°	15000	28	80	108
Geosynch	$\sim 3 \times 10^6$	300	250	550

\*\* From J. W. Watts and J. J. Wright, NASA Marshall Space Flight Center

<sup>†</sup> Cycle 21, 0.99 confidence, Solar Maximum



shielding might apply to a typical internal point. Comparison of Figure 1 and Table 1 shows that at all altitudes and inclinations there are some restrictions imposed by the radiation environment on the application of certain materials and components. For the higher inclinations and altitudes manned operations will be limited in duration by the radiation exposure limits, and at the higher altitudes extravehicular activity in space suits may be prohibited.

The space radiation environment of significance in this discussion, up to geosynchronous altitudes, consists of three components: the galactic cosmic radiation, the trapped radiation belts, and solar flare particles. All three components are influenced to some degree by solar activity. Many aspects of the particle radiation environment and temporal changes caused by solar activity are covered in Working Group B3 of the Solar-Terrestrial Predictions Workshop.

The galactic cosmic radiation consists of the nuclei of the elements and electrons. A recent description of the characteristics of the flux may be found in Meyer *et al.* (1974). The nuclei are predominant in number above a given energy and deposit essentially all of the radiation dose. The cosmic ray nuclei reaching Earth have energies greater than  $\sim 30$  MeV/nucleon, and the mean particle energy is near GeV. Thus, the particles are very penetrating, and spacecraft shielding is not very effective in reducing dose. In all orbits,  $\sim 5$  to 10 percent of the total dose (in rads) within typical spacecraft shielding is due to cosmic rays. Because of the unique effects of the highly ionizing heavy nuclei in cosmic rays, they often must be considered separately.

The part of the cosmic ray energy spectrum below  $\sim 2$  GeV/nucleon is influenced by solar activity via interplanetary magnetic fields carried there by plasma from the Sun. This modulation of the galactic spectrum creates a maximum in the differential energy spectrum of cosmic rays between 200 and 800 MeV/nucleon, depending upon solar activity. Similarly, the integral number of the particles is affected as a function of solar activity. Although other short-term variations in galactic cosmic ray intensity occur, the variation with the 11-year solar cycle is the only one of significance in radiation effects considerations.

This variation would cause approximately a factor of 2 variations in cosmic ray dose within spacecraft at geosynchronous orbits and essentially none in low-Earth, low-inclination orbits. The radiation dose in spacecraft structures due to cosmic rays has been treated by Curtis and Wilkinson (1971) and Burrell and Wright (1972). The influence of secondary heavy nuclei has received some consideration, but not large cosmic ray showers.

The trapped belt radiation has received extensive attention in the literature. See, for example, the "Trapped Radiation Handbook," Cladis (1977). In terms of solar-induced variations there are two general cases of significance in radiation effects prediction. The first is the influence of solar cycle activity on the inner belt proton populations, which has been summarized by Chan *et al.* (1977). This variation in particle population will produce long-term (11-year) variations of about a factor of 2 in radiation dose rate within spacecraft for low Earth orbits. It is to be noted that uncertainty in

predicted dose rates is of the same order, especially in low Earth orbits where the gradient of particle flux with altitudes is steep.

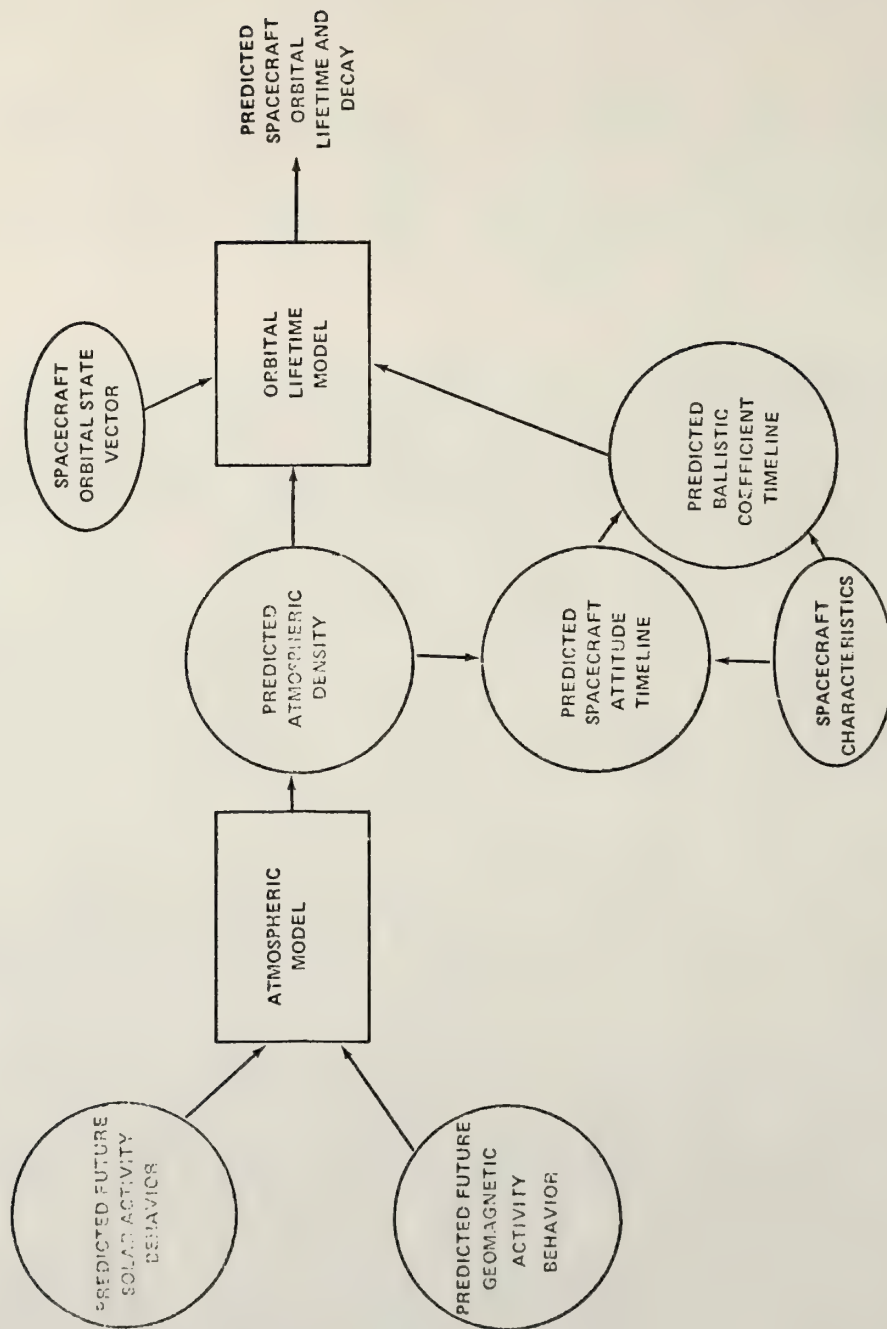
The second general case of variation is that of the outer belt electrons which produce most of the radiation dose at geosynchronous orbits. The electrons produce radiation dose directly at thin shielding thickness ( $<1 \text{ gm/cm}^2$ ) and via bremsstrahlung X-rays at greater thicknesses. The bremsstrahlung dose is highly dependent on the shielding material. In addition to long-term (11-year) variations, the outer belt electrons undergo order of magnitude changes in flux over time scales of weeks. As a result, radiation dose calculations for long periods use models of the flux which average over these variations (Vette and Singley, 1972). If dose rates become significant for short periods (e.g., manned missions to geosynchronous orbits), a probability distribution method for predicting dose for the outer belt electrons will need to be developed.

When predicting radiation effects due to solar flares, the temporal variations are a dominant feature. Solar flare radiation is a significant hazard for orbital inclinations greater than  $\sim 50$  degrees and altitudes above a few Earth radii. The individual solar particle events vary in particle constituents, energy spectra, and particle fluence per flare. Usually a few very large flares dominate the total particle fluence for a solar cycle. Because of the quasirandom nature of the flares, radiation doses from flares are expressed in terms of probability models. Burrell (1971) developed such a model and performed sample dose calculations based upon data of the 19th solar cycle. King (1974) has summarized data from the 19th and 20th cycles and extended Burrell's statistical model to predict solar proton fluence probabilities for the 21st cycle. The data of Table 2 indicate that solar flares present significant radiation problems, especially for manned flight beyond low Earth orbit.

#### SOLAR PREDICTIONS AND SPACECRAFT ORBITAL LIFETIME

A semi-analytical method is used in most spacecraft orbital lifetime prediction models to estimate the decay history and the lifetime of a near-Earth orbiting spacecraft perturbed by atmospheric drag. For most near-Earth orbits with small eccentricity, the perturbations due to other forces (i.e., solar-lunar gravity perturbations, solar radiation pressure, and electromagnetic effects) are overshadowed by the effects caused by uncertainties in the calculation of atmospheric drag. For this reason, efforts to incorporate additional perturbing forces are often unwarranted. The approach used to estimate the orbital decay usually adopts a combination of general and special perturbation techniques so that the analysis preserves sufficient rigor to insure accuracy and adequate numerical emphasis to include a rather sophisticated atmospheric density model in an efficient simulation. Basically the procedure is to extend a system of ordinary differential equations for a set of well-defined mean orbital elements which describe the complete motion of a spacecraft about an oblate Earth to include numerically the drag effect due to a rotating atmosphere. The program is designed with the objective that it can provide a means to estimate, with reasonable accuracy, the orbital decay history and the orbital lifetimes efficiently and quickly. Figure 2 illustrates the principal inputs and components of a spacecraft orbital lifetime and decay prediction procedure.





NASA/MSFC/ES-81

FIGURE 2. Solar predictions and spacecraft orbital lifetime.

A major difficulty in predicting the orbital lifetime of an orbiting spacecraft results because the future characteristics of the atmospheric density are not deterministically known. This makes it necessary to specify orbital lifetimes in a probabilistic manner. Comparisons of predicted spacecraft decay versus actual observed decay reveal deviations which can be attributed to an inadequate deterministic atmospheric model, noisy tracking data, or deviations in the stochastic variables associated with the lifetime prediction problem (i.e., ballistic coefficient, solar flux, and geomagnetic activity). The ballistic coefficient is a function of the spacecraft mass, drag coefficient, and effective cross section area. It is observed to vary with the spacecraft orientation and flight region. In addition, predictions of the solar flux and geomagnetic activity values, over either short or long periods of time, are available only in terms of statistical predictions with significant uncertainties (Liu, 1975).

Straus and Hickman (1978) have described the characteristics of several widely used atmospheric density models and have reviewed studies in which the predictions of these models have been compared with observational data. They have also assessed the relative advantages and limitations of the models in current use. They conclude that models produced prior to 1970 were developed from data bases with significant limitations. The overall accuracy of the recent models is summarized as, on the average, as good as the measurements of atmospheric density and composition. Individual calculations of the ratio for measured to model values show a scatter (standard deviation) of approximately 15 to 20 percent for the total density. For long-term spacecraft orbital lifetime estimates a model with a correct mean density averaged over several months is quite adequate even if the errors averaged over a single orbit are quite large.

The upper atmosphere density is strongly influenced by the changing levels of solar activity. This, in turn, directly affects the spacecraft drag and orbital lifetime. It is the ultraviolet solar radiation that heats and causes structural changes in the Earth's upper atmosphere. One component of this radiation relates to the active regions on the solar disk and varies from day to day; whereas the other component relates to the solar disk itself and varies more slowly with the 11-year solar cycle. The atmosphere reacts differently to each of these two components (NASA, 1973).

The 10.7 cm solar flux is generally used as a readily available index of solar ultraviolet radiation. It also consists of a disk component and an active area component. When the 10.7 cm flux increases, there is an increase in the upper atmosphere density. For a given increase in the disk component of the 10.7 cm flux, however, the density increases much more than for the same increase in the active area component. For all practical purposes the active area component is linearly related to the daily 10.7 cm flux and the disk component to the 10.7 cm flux averaged over a few solar rotations (i.e., six is recommended by Jacchia (1977)). Therefore, the need exists for accurate predictions of the 10.7 cm solar flux, or preferably the ultraviolet radiation, for periods of a few weeks up to 10 or 15 years in the future in order to compute the expected atmospheric density values for use in orbital lifetime prediction models. The orbital lifetime is inversely related to the upper atmosphere density.



An error, for example, in the prediction of the peak in the average solar activity on the ascending slope of the solar cycle by +30 percent, for a spacecraft at approximately 400 km altitude and having a nominal predicted lifetime of approximately 20 months, produces a decrease in the lifetime by approximately 30 percent. While this linear error relationship does not hold for all combinations of variations in solar activity, orbital altitudes, and ballistic factors, the example does illustrate the importance in the development of either a deterministic long-term solar activity prediction procedure or a statistical procedure with much closer confidence (error) bounds than now exists.

## NEEDS FOR SOLAR ACTIVITY PREDICTIONS

For a given spacecraft configuration, the types of operational anomalies that can occur and their probabilities of occurrence for a given charge buildup-discharge event probably can be defined with reasonable accuracy. If one had the capability to predict the probability of an encounter with solar-induced environmental conditions that could result in anomalous operations, it should be possible to develop operational protocols which control the deleterious effects of an anomaly without adding unwarranted complexity to the system and without unnecessarily reducing operational capability during environmental quiet periods. Therefore, a predictive capability for solar behavior appears to be a goal well worth pursuing relative to spacecraft charging interactions, operational planning, and mission analyses.

In terms of improved predictions in the solar-terrestrial environment that apply to particle radiation effects on a spacecraft, it is obvious that immediate design and operational benefits would result from improved predictions of solar flare activity. This is certainly true for long-term radiation predictions required for assessment of effects on materials and electronic components. Because single solar particle events can produce a large fraction of the total 11-year cycle fluence, short-term predictions and warning methods would be essential for manned operations in geosynchronous orbit. Also, for geosynchronous orbits a better understanding of solar activity influences on the outer belt electron populations would be valuable for any long-term predictions as well as perhaps improving any models for short-term radiation dose probability calculations. It is believed, within the accuracy to which radiation effects can now be estimated, that our present knowledge of the long-term (11-year cycle) variations is adequate for both cosmic ray dose and fluence calculations and for dose calculations related to inner belt protons.

There continues to exist a critical need for more accurate predictions of short- and long-term solar activity to use in atmospheric density models. This will be required not only for the monitoring and accurate estimation of the orbital lifetime and decay for the large numbers of spacecraft and "junk" now in orbit but for the more economical and efficient estimation of future spacecraft missions, especially in the near-Earth orbital environment. The expected future state of the upper atmosphere plays an important part in the decisions associated with spacecraft missions. The critical dependence on solar activity predictions can easily be seen by looking at the flow chart shown in Figure 2.

## REFERENCES

- Baughner, Charles R. (1978): Private communication, NASA Marshall Space Flight Center, Huntsville, Alabama.
- Benton, E. V., D. D. Peterson, J. V. Bailey, and T. A. Parnell (1977): Health Physics, 32:15-19.
- Binder, D., E. C. Smith, and A. B. Holman (1975): IEEE Transactions on Nuclear Science, NS-22, 6:2675-2680.
- Burrell, M. O. (1971): The risk of solar proton events to space missions, NASA Tech. Note TN D-6379, National Aeronautics and Space Administration, Washington, D. C.
- Burrell, M. O., J. J. Wright, and J. J. Watts (1968): Analysis of energetic space radiation and dose rates, NASA Tech. Note TN D-4404, National Aeronautics and Space Administration, Washington, D. C.
- Burrell, M. O., and J. J. Wright (1972): The estimation of galactic cosmic ray penetration and dose rates, NASA Tech. Note TN D-6600, National Aeronautics and Space Administration, Washington, D. C.
- Chan, K. W., D. M. Sawyer, and J. I. Vette (1977): Trapped radiation population. In: Trapped Radiation Handbook, DNA 2524H, Change 5, Defense Nuclear Agency.
- Cladis, J. B., Editor (1977): Trapped Radiation Handbook, Defense Nuclear Agency, DNA 2524H, Change 5.
- Curtis, S. B., and M. C. Wilkinson (1971): Galactic cosmic ray heavy primary secondary doses. In: NASA Tech. Memo. TM X-2440, National Aeronautics and Space Administration, Washington, D. C.
- Davis, A. H., and I. Harris (1961): Interaction of a charged satellite with the ionosphere. In: Rarefied Gas Dynamics, L. Talbot, Editor, New York, Academic Press.
- DeForest, Sherman E. (1977): The plasma environment at geosynchronous orbit. In: Proceedings of the Spacecraft Charging Technology Conference, C. P. Pike and R. R. Lovell, Editors, NASA Tech. Memo. TM X-73537, 37-52.
- Fishman, G. J., and R. W. Austin (1976): Nuclear Instruments and Methods, 140:193-196.
- Fredricks, R. W., and F. L. Scarf (1973): Observations of spacecraft charging effects in energetic plasma regions. In: Photon and Particle Interactions with Surfaces in Space, R. V. L. Grard, Editor, Dordrecht-Holland, D. Reidel Publishing Company, 277-308.
- Hill, C. W. (1978): Survey of radiation effect. Report, Contract NAS8-32387, NASA Marshall Space Flight Center, Huntsville, Alabama.



- Inouye, George T. (1977): Spacecraft charging anomalies on the DSCS 11, launch 2 satellites. In: Proceedings of the Spacecraft Charging Technology Conference, C. P. Pike and R. R. Lovell, Editors, NASA Tech. Memo. TM X-73537, 829-852.
- Jacchia, L. G. (1977): Thermospheric temperature, density and composition: new models. Special Report 375, Smithsonian Astrophysical Observatory, Cambridge, Massachusetts.
- King, J. H. (1974): Journal of Spacecraft, 11, 6:401-408.
- Liu, J. J. F. (1975): Orbital decay and lifetime estimation. Report TN-240-1441, Northrop Services, Inc., Huntsville, Alabama.
- May, T. C., and M. H. Woods (1978): A new physical mechanism for soft errors in dynamic memories. Intel Corporation, Santa Clara, California.
- McPherson, D. A., D. P. Cauffman, and W. Schober (1975): Spacecraft charging at high altitudes: SCATHA satellite program. J. Spacecraft and Rockets, 12:621.
- Meyer, P., R. Ramaty, and W. R. Webber (1974): Physics Today, October:23-32.
- Napoli, Joseph, and Joseph Seliga (1977): RCA Satcom 3-Axis Spacecraft experience at geosynchronous altitude. In: Proceedings of the Spacecraft Charging Technology Conference, C. P. Pike and R. R. Lovell, Editors, NASA Tech. Memo. TM X-73537, 865-871.
- NASA Vehicle Design Criteria (1973): Models of Earth's atmosphere (90 to 2500 km). NASA SP-8021, National Aeronautics and Space Administration, Washington, D. C.
- Parker, L. W. (1976): Computation of collisionless steady-state plasma flow past a charged disk. NASA CR-144159, Lee W. Parker, Inc., Concord, Massachusetts.
- Parker, L. W., and E. C. Whipple (1970): Theory of spacecraft sheath structure, potential and velocity effects on ion measurements by traps and mass spectrometers. J. Geophys. Res., 75:4720.
- Reasoner, D. L., W. Lennartsson, and C. R. Chappell (1976): The relationship between ATS-6 spacecraft charging occurrences and warm plasma encounters. In: Spacecraft Charging by Magnetospheric Plasmas, Progress in Astronautics and Aeronautics, Vol. 47, New York, American Institute of Aeronautics and Astronautics, 89-101.
- Robbins, A., and C. D. Short (1977): Space environmental effects on the Skynet 2B spacecraft. In: Proceedings of the Spacecraft Charging Technology Conference, C. P. Pike and R. R. Lovell, Editors, NASA Tech. Memo. TM X-73537, 853-863.
- Singer, S. F., and E. H. Walker (1962): Photoelectric screening of bodies in space. Icarus, 1:7.

- Straus, Joe M., and David R. Hickman (1978): Predictability of upper atmospheric density and composition. Aerospace Report No. SSL-78 (9553)-2, The Aerospace Corporation, Los Angeles, California.
- Vette, J. I., and G. W. Singley (1972): Model environment for outer zone electrons. NSSDC 72-13, National Space Science Data Center.
- Warman, E. A., Editor (1971): Proceedings of the National Symposium on Natural and Manmade Radiation in Space, NASA TM X-2440, National Aeronautics and Space Administration, Washington, D. C.
- Watts, J. W., Jr., and J. J. Wright (1976): Charged particle radiation environment for the Spacelab and other missions in low Earth orbit, Revision A. NASA Tech. Memo. TM X-73358, National Aeronautics and Space Administration, Washington, D. C.
- Whipple, E. C., O. M. Warnock, and R. H. Winkler (1974): Effect of satellite potential on direct ion measurements through the plasmopause. J. Geophys. Res., 79:179.



PREDICTION OF SPACECRAFT POTENTIALS  
AT GEOSYNCHRONOUS ORBIT

H.B. Garrett, A.G. Rubin, and  
C. P. Pike  
Space Physics Division  
Air Force Geophysics Laboratory  
Hanscom AFB MA 01731

The Air Force Geophysics Laboratory has developed several levels of models for the computation of spacecraft potentials. In parallel with these efforts a systematic study of the state of the low energy plasma environment has been under way. In this paper we will present the synthesis of these efforts into a consistent picture of the level of charging to be expected on a "typical" geosynchronous satellite in eclipse as a function of geomagnetic activity. As such results also allow the prediction of the maximum differential potentials to be expected between electrically isolated, illuminated and shadowed surfaces on a space vehicle, they are of great value to the spacecraft designer.

INTRODUCTION

This paper outlines two relatively straightforward techniques for determining spacecraft potentials in the limit of a "thick sheath" surrounding the spacecraft. A statistical model of the various features of the geosynchronous environment based on ATS-5 and ATS-6 data and an analytic model capable of detailed simulation of the low energy geosynchronous environment are also discussed. The results from these two environmental models are then combined with the charging models in order to provide estimates of the relationships between the geomagnetic  $K_p$  index and spacecraft potential. The results are compared with actual potential measurements from ATS-5 and ATS-6.

CHARGING MODELS

No matter what form the calculation of spacecraft potential takes, the basic equation is one of current balance:

$$J_E - (J_I + J_{SE} + J_{SI} + J_{BSE} + J_{PH}) = 0 \quad (1)$$

where:  $J_E$  = incident electron current  
 $J_I$  = incident ion current  
 $J_{SI}$  = secondary electron current due to ions

$J_{SE}$  = Secondary electron current due to electrons  
 $J_{BSE}$  = Backscattered electron current  
 $J_{PH}$  = Photoelectron current

All of these currents are complex functions of spacecraft potential, the spacecraft sheath, and (except for  $J_{PH}$ ) the ambient low energy plasma. A variety of detailed codes have been developed to handle equation 1 (Whipple, 1965; Rothwell et al., 1977; Laframboise and Prokopenko, 1977; and Parker, 1977). Although capable of a fairly accurate treatment in both time and space, these models require large amounts of computer time or do not include all of the various currents in equation 1. We will instead present two simple models based on the assumption of a "thick sheath" (that is, the region over which the spacecraft effects the ambient plasma is much larger than the spacecraft). As an example of the validity of this assumption at geosynchronous orbit, the sheath thickness, approximated by the Debye length, is  $\sim 250$  M (for a density of  $1/\text{cm}^3$  and a temperature of 1 KeV).

The distribution function  $f$  for the ambient particles is related to the distribution function  $f'$  at the surface of the spacecraft by the relationship (for an isotropic distribution):

$$f(E) = f'(E') \quad (2)$$

$$\text{and} \quad E = E' + qV \quad (3)$$

where:  $E$  = Energy of ambient particles  
 $q$  = Particle charge  
 $V$  = Satellite potential

In the simple case of a Maxwellian distribution this reduces for the thick sheath approximation to:

$$J_E = J_{EO} e^{-|qV|/kT_e} \quad V < 0 \quad (4)$$

$$= J_{EO} (1 + |qV|/kT_e) \quad V > 0$$

$$\text{and} \quad J_I = J_{IO} e^{-|qV|/kT_I} \quad V > 0 \quad (5)$$

$$= J_{IO} (1 + |qV|/kT_I) \quad V < 0$$

where:  $J_{EO}$  = Ambient electron current  
 $J_{IO}$  = Ambient ion current  
 $k$  = Boltzmann constant  
 $T_e$  = Ambient electron temperature  
 $T_I$  = Ambient ion temperature

In Garrett and Rubin (1978), it was assumed (based on more sophisticated calculations) that

$$\begin{aligned}
 J_{SE} &\sim a J_E \\
 J_{BSE} &\sim b J_E \\
 J_{SI} &\sim c J_I
 \end{aligned} \quad (6)$$



where:

$$\begin{aligned} a &\sim .4 \\ b &\sim .2 \\ c &\sim 3 \end{aligned}$$

This means that equation 1 becomes, for a negatively charged spacecraft:

$$(1-a-b) J_{E0} e^{-|qV|/kT_e} - (1+c) J_{I0} (1 + \frac{|qV|}{T_I}) - J_{PH} = 0 \quad (7)$$

If the satellite is in eclipse,  $J_{PH}=0$ . Further, under most conditions at geosynchronous orbit,  $|qV_0|/T_I$  is generally less than 1 so that equation 7 reduces to:

$$|qV_0| \sim -kT_e \ln(J_{E0}/10J_{I0}) \quad (8)$$

Results from ATS-5 and ATS-6 are plotted in Figure 1 for two different definitions of the temperature:  $T(AVG)$  and  $T(RMS)$  (see discussion later in text).

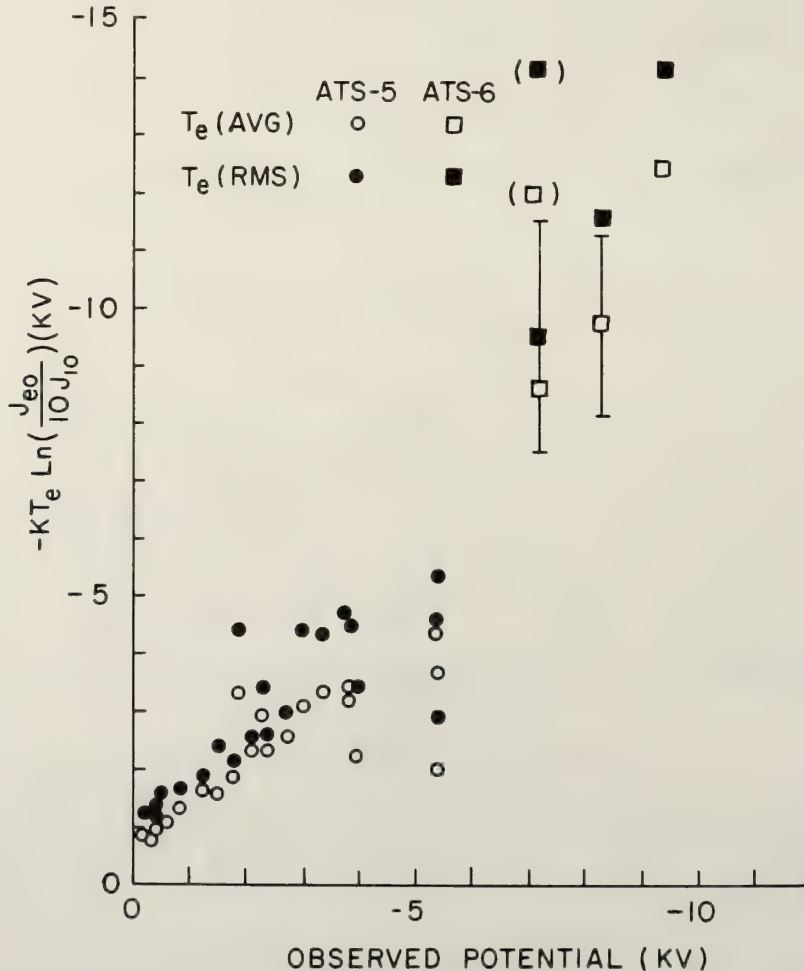


FIGURE 1. Satellite potential as estimated by equation 8 versus the potential actually observed by ATS-5 and ATS-6, the marked values are believed to have occurred during rapid changes in the ambient plasma.

The standard deviations of the predicted from the observed voltages are  $\pm 1700V(T(AVG))$  and  $\pm 2000 V(T(RMS))$ . (Note: the ATS-6 values that are marked occurred during rapid plasma variations and, if ignored, would lower the standard deviations). Figure 1 not only demonstrates the validity of equation 8, but implies the existence of a threshold electron temperature above which the satellite potential is linear (i.e., about 1kV). Such behavior is due to the fact that at some intermediate energy, the secondary yield is greater than 1 (Rubin et al., 1978). The electron temperature must be several times greater than this threshold energy before significant charge buildup occurs.

If, as in the case of ATS-5 and ATS-6, we have available the actual particle distribution functions we can significantly improve on equations 4,5, and 6 by using them in the calculation of the currents of equation 1. As discussed in Garrett (1978), the currents in equation 1 are each carefully computed by integrating the observed distribution functions as outlined in Whipple (1965) and DeForest (1972). As these are all functions of satellite potential, the potential is varied until equation 1 holds.

The results of these more detailed calculations are presented in Figure 2 for two cases: for the ambient spectra measured immediately before (or immediately after) the eclipse interval (sunlit) and during eclipse (eclipsed). The potential in eclipse is accurately predicted within  $\pm 700$  volts by this model (note: the greater uncertainty in the

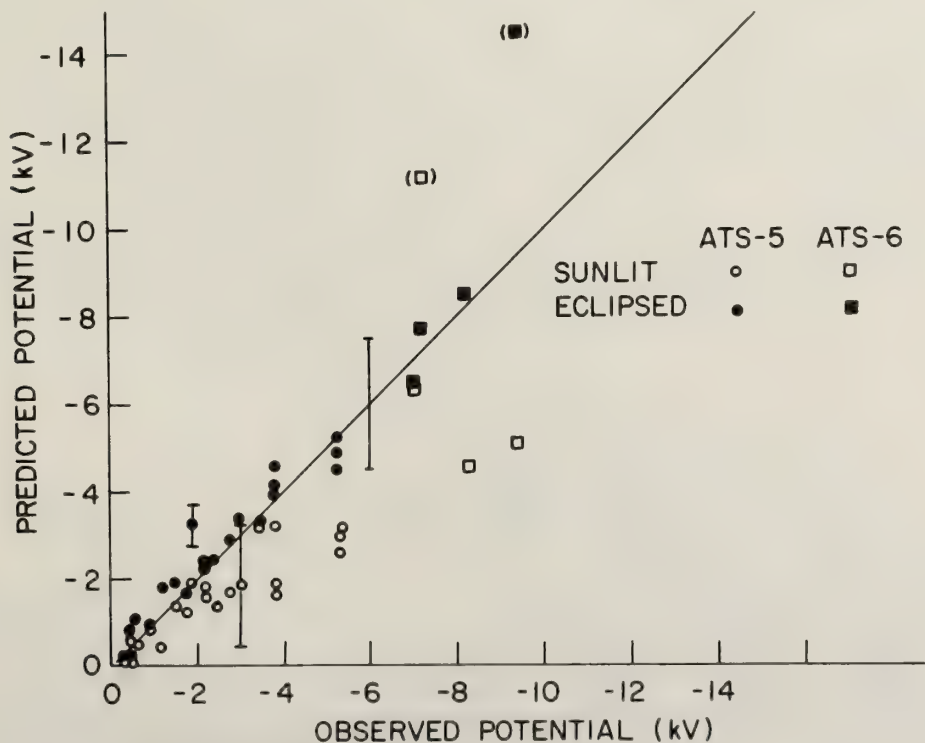


FIGURE 2. The potential as predicted by the model of Garrett (1978) versus the observed ATS-5 and ATS-6 potentials. The marked values occurred during periods of rapidly changing potential.



sunlit spectra is believed due to the measurement techniques employed and not the model).

From the preceding we can see that even a very simple model can make reasonable predictions of the potential on a spacecraft in eclipse. Such information is important as it provides a good quantitative estimate of the maximum potentials likely to develop between isolated surfaces on a space vehicle. In subsequent sections, after presenting models of the ambient environment we will return to the two models presented here in order to estimate the relationship between satellite potential in eclipse and geomagnetic activity.

## THE GEOSYNCHRONOUS ENVIRONMENT

A variety of models exist which describe the low energy (0-100keV) near-earth space environment (see review by Garrett, 1979). AFGL has developed two descriptions of the environment. The first of these is a statistical compilation of the occurrence frequency of various parameters at geosynchronous orbit as functions of local time and geomagnetic activity. The second model is a set of analytic expressions capable of predicting simultaneous variations in the geosynchronous plasma as functions of local time and geomagnetic activity. This latter model, though based on a relatively limited data base, will allow an estimation of how potential varies with geomagnetic activity and, simultaneously, local time.

The first model we want to discuss is based on a statistical compilation of 50 days of data from the ATS-5 and ATS-6 satellites. Very briefly, the 4 moments of the observed distribution function were calculated from the data base as follows:

$$\langle N_i \rangle = 4\pi \int_0^\infty (V^0) f_i V^2 dV = n_i \quad (9)$$

$$\langle NF_i \rangle = \int_0^\infty (V^1) f_i V^2 dV = \frac{n_i}{2\pi} \left[ \frac{2kT_i}{\pi m_i} \right]^{1/2} \quad (10)$$

$$\langle P_i \rangle = 4\pi (1/3 m_i) \int_0^\infty (V^2) f_i V^2 dV = n_i kT_i \quad (11)$$

$$\langle EF_i \rangle = (1/2 m_i) \int_0^\infty (V^3) f_i V^2 dV = \frac{m_i n_i}{2} \left[ \frac{2kT_i}{\pi m_i} \right]^{3/2} \quad (12)$$

where:

- $\langle N_i \rangle =$  number density for species i (number/cm<sup>3</sup>)
- $\langle NF_i \rangle =$  number flux for species i (number/cm<sup>2</sup>sec-sr)
- $\langle P_i \rangle =$  pressure for species i (dynes/cm<sup>2</sup>)

$\langle EF_i \rangle =$  energy flux for species  $i$  (ergs/cm<sup>2</sup>sec-sr)

The integral results on the right are for a Maxwell-Boltzmann distribution:

$$f_i(V) = n_i \left[ \frac{m_i}{2\pi kT_i} \right]^{3/2} e^{-m_i V_i^2 / 2kT_i} \quad (13)$$

where:  $n_i$  = number density of species  $i$

$m_i$  = mass of species  $i$

$T_i$  = temperature of species  $i$

$V_i$  = velocity of species  $i$

$k$  = Boltzmann constant

$f$  = distribution function

The description of the plasma in terms of these quantities is quite useful as not only are they physically meaningful in their own right, but they can be used to derive a Maxwellian or a two-Maxwellian distribution of the environment (see Garrett and DeForest, 1979). As an example, in Figure 3, we have plotted the occurrence frequency of the ambient current (the number flux multiplied by  $\pi q$ ) and two estimates of the temperature (Garrett et al., 1978). Before discussing these results in any more detail, an important point must be considered in the estimation of the plasma temperature from the four moments of the distribution function.

A single temperature cannot be defined if the plasma is not Maxwellian or if the plasma consists of two or more Maxwellian components - circumstances which occur frequently at geosynchronous orbit. As a test of this effect, we have defined two "temperatures".

$$T \text{ (AVG)} = \frac{\langle P \rangle}{\langle N \rangle} \quad (14)$$

$$T \text{ (RMS)} = \frac{\langle EF \rangle}{2\langle NF \rangle} \quad (15)$$

These temperatures will be equal and have meaning as temperatures if and only if the plasma is a Maxwellian plasma (i.e., representable by equation 13). The marked difference between  $T(\text{AVG})$  and  $T(\text{RMS})$  in Figure 3 is a direct result of the absence of a Maxwellian plasma at geosynchronous orbit.

If the plasma is considered to consist of two Maxwellian components, then we can define two temperatures and two densities as follows for species  $i$ :

$$f_{2i}(V_i) = n_{1i} \left[ \frac{m_i}{2\pi kT_{1i}} \right]^{3/2} e^{-m_i V_i^2 / 2kT_{1i}} + n_{2i} \left[ \frac{m_i}{2\pi kT_{2i}} \right]^{3/2} e^{-m_i V_i^2 / 2kT_{2i}} \quad (16)$$



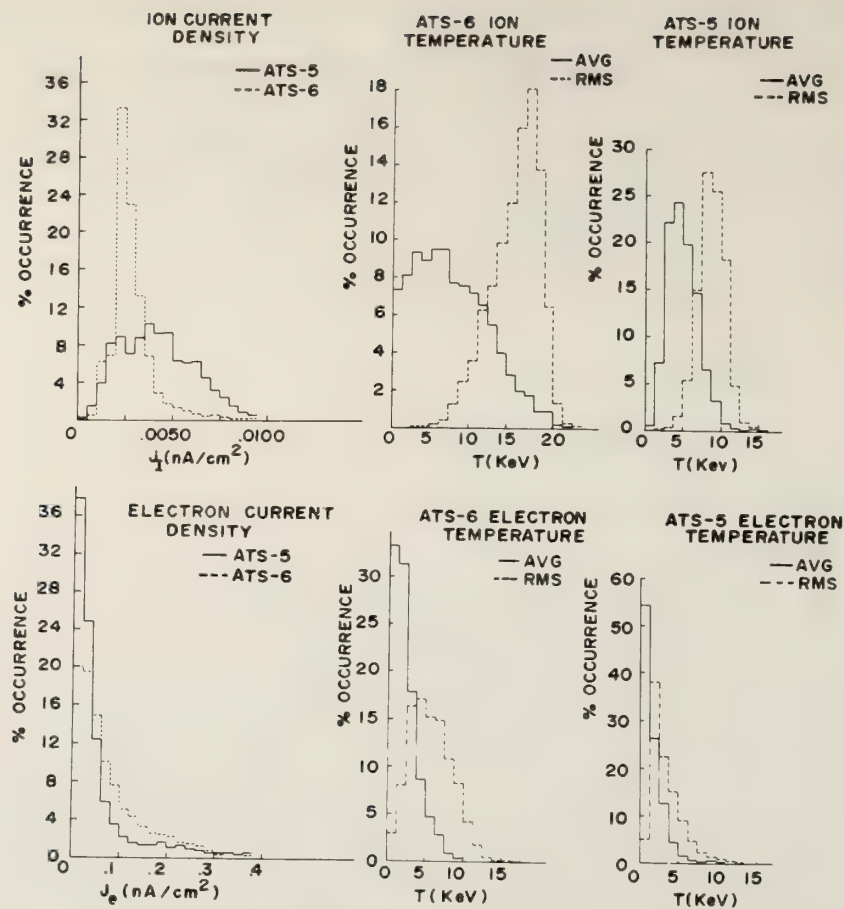


FIGURE 3. Statistical occurrence frequencies of ATS-5 (1969 and 1970 data) and ATS-6 (1974 and 1976 data) electron and ion current densities, T(RMS), and T(AVG). The ATS-6 values should be considered provisional.

where  $n_{1i}$ ,  $T_{1i}$ ,  $n_{2i}$ , and  $T_{2i}$  can be derived directly from equations 9 through 12. T(AVG) and T(RMS) in terms of these quantities are:

$$T \text{ (AVG)} = k \frac{n_1 T_1 + n_2 T_2}{n_1 + n_2} \quad (17)$$

$$T \text{ (RMS)} = k \frac{n_1 T_1^{3/2} + n_2 T_2^{3/2}}{n_1 T_1^{1/2} + n_2 T_2^{1/2}} \quad (18)$$

For typical values we find:

For Electrons

$$\begin{aligned} n_1 &= 1/\text{cm}^3 \\ T_1 &= 500 \text{ eV} \\ n_2 &= 0.2/\text{cm}^3 \\ T_2 &= 6000 \text{ eV} \\ T(\text{AVG}) &\approx 1400 \text{ eV} \\ T(\text{RMS}) &\approx 2750 \text{ eV} \end{aligned}$$

For Ions

$$\begin{aligned} n_1 &= 0.6/\text{cm}^3 \\ T_1 &= 100 \text{ eV} \\ n_2 &= 0.6/\text{cm}^3 \\ T_2 &= 9000 \text{ eV} \\ T(\text{AVG}) &\approx 4550 \text{ eV} \\ T(\text{RMS}) &\approx 8150 \text{ eV} \end{aligned}$$

These values are very close to the average values for ATS-5 shown in Figure 3, and we believe readily explain the differences between  $T(\text{AVG})$  and  $T(\text{RMS})$ . It is also important to note that  $T_1$  and  $T_2$  are not necessarily valid temperatures. They are the result of a definite fitting process - their primary use being as scaling parameters in obtaining a two Maxwellian fit to the distribution function. A common problem is that when the plasma is close to Maxwellian, one of the temperatures will be unrealistically large even though the fitted distribution is quite close to the actual one.

Our second type of model is what we have termed an analytic simulation. What a statistical model, such as just presented, fails to do is to maintain the proper time relationship between the various parameters (that is, if A goes up, does B go up or down?). In order to maintain the correlated variations in different parameters, we have made use of linear regression techniques. Three hour averages of the four moments of the electron and ion distribution functions for ten carefully selected days \* of ATS-5 were fit by linear regression techniques to an equation varying linearly in the geomagnetic index  $A_p$  and diurnally and semidiurnally in local time (LT):

$$\begin{aligned} M(\text{LT}, A_p) &= (a_0 + a_1 A_p) (b_0 + b_1 \cos(\frac{\pi}{12} (\text{LT} + t_1))) \\ &\quad + b_2 \cos(\frac{\pi}{6} (\text{LT} + t_2)) \end{aligned}$$

---

\* The days were carefully selected to correspond to periods when plasma was injected while the satellite was at midnight - periods when maximum spacecraft charging is believed to take place.



# ELECTRON TIME VARIATIONS

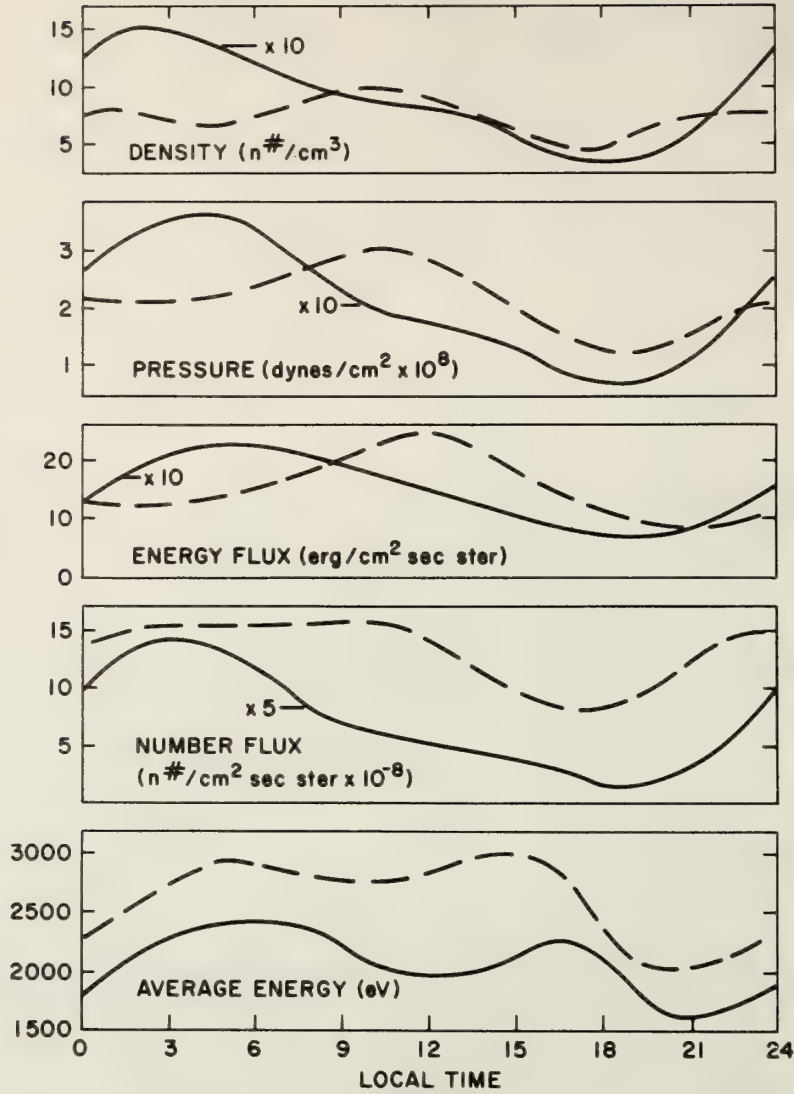


FIGURE 4a. Variations in the 4 moments and the mean energy of the electrons according to the AFGL simulation code for  $A_p=15$  (solid line) and  $A_p=207$  (dashed line).

where:  $M(LT, A_p)$  = predicted value of the moment,  $M$ , at local time,  $LT$ , and

$$\text{for activity level } A_p = \frac{\sum a_p}{8} \quad (\text{i.e., daily average of } a_p)$$

$a_0, a_1, b_0, b_1, b_2, t_1, t_2$  = coefficients determined by the regression (see Table 1).

TABLE 1. MULTIPLE REGRESSION COEFFICIENTS \*

	$C_0$	$C_1$	$C_2$	$C_3$	$C_4$	$C_5$	$C_6$	$C_7$	$C_8$	$C_9$
1	.38 + 02	-.42 + 02	.22 + 02	-.21 + 01	-.23 + 02	.43 + 00	-.73 - 01	-.60 - 01	-.80 - 01	.44 - 01
2	.99 + 02	-.27 + 02	.23 + 01	-.28 + 01	-.33 + 01	.21 + 00	.98 - 02	-.35 - 01	-.48 - 01	.37 - 01
3	.52 + 01	-.10 + 02	.64 + 01	.26 + 01	-.48 + 01	.13 + 00	-.28 - 01	-.33 - 01	-.20 - 01	.23 - 02
4	.77 + 02	-.45 + 01	.16 + 02	-.81 + 00	-.31 + 01	.16 + 00	-.49 - 01	.42 - 01	-.64 - 01	.24 - 01
5	.36 + 02	-.74 + 02	.50 + 02	.21 + 02	-.11 + 02	.91 + 00	-.33 - 01	-.51 + 00	-.17 + 00	-.47 - 01
6	.19 + 02	-.52 + 00	.31 + 01	-.63 - 01	-.50 + 00	.38 - 01	-.12 - 01	.15 - 01	-.18 - 01	.34 - 02
7	.47 + 02	-.68 + 02	.38 + 02	.76 + 01	-.44 + 02	.81 + 00	-.24 + 00	-.45 - 01	-.11 + 00	.50 - 01
8	.62 + 01	-.95 + 00	.14 + 01	-.13 + 00	-.30 + 00	.15 - 01	-.34 - 02	.76 - 03	-.47 - 02	.25 - 02

where:

- $I = 1 \ n_e \times 100 \text{ (number/cm}^3\text{)}$      $5 \ E F_e \times 100 \text{ (erg/cm}^2\text{-sec-ster)}$   
 $2 \ n_I \times 100 \text{ (number/cm}^3\text{)}$      $6 \ E F_I \times 100 \text{ (erg/cm}^2\text{-sec-ster)}$   
 $3 \ P_e \times 10^{10} \text{ (dynes/cm}^2\text{)}$      $7 \ N F_e \times 10^{-6} \text{ (number/cm}^2\text{-sec-ster)}$   
 $4 \ P_I \times 10^{10} \text{ (dynes/cm}^2\text{)}$      $8 \ N F_I \times 10^{-6} \text{ (number/cm}^2\text{-sec-ster)}$

such that:

$$M_I(A_p, Lt) = C_0 + C_1 \cos\left(\frac{2\pi}{24} t\right) + C_2 \sin\left(\frac{2\pi}{24} t\right) + C_3 \cos\left(\frac{4\pi}{24} t\right) + C_4 \sin\left(\frac{4\pi}{24} t\right) + C_5 A_p' + C_6 A_p' \cos\left(\frac{2\pi}{24} t\right) + C_7 A_p' \sin\left(\frac{2\pi}{24} t\right) + C_8 A_p' \cos\left(\frac{4\pi}{24} t\right) + C_9 A_p' \sin\left(\frac{4\pi}{24} t\right)$$

$$t = Lt + 6.5 \quad A_p' = 8 \cdot A_p$$

Lt = local time

\*second number is power



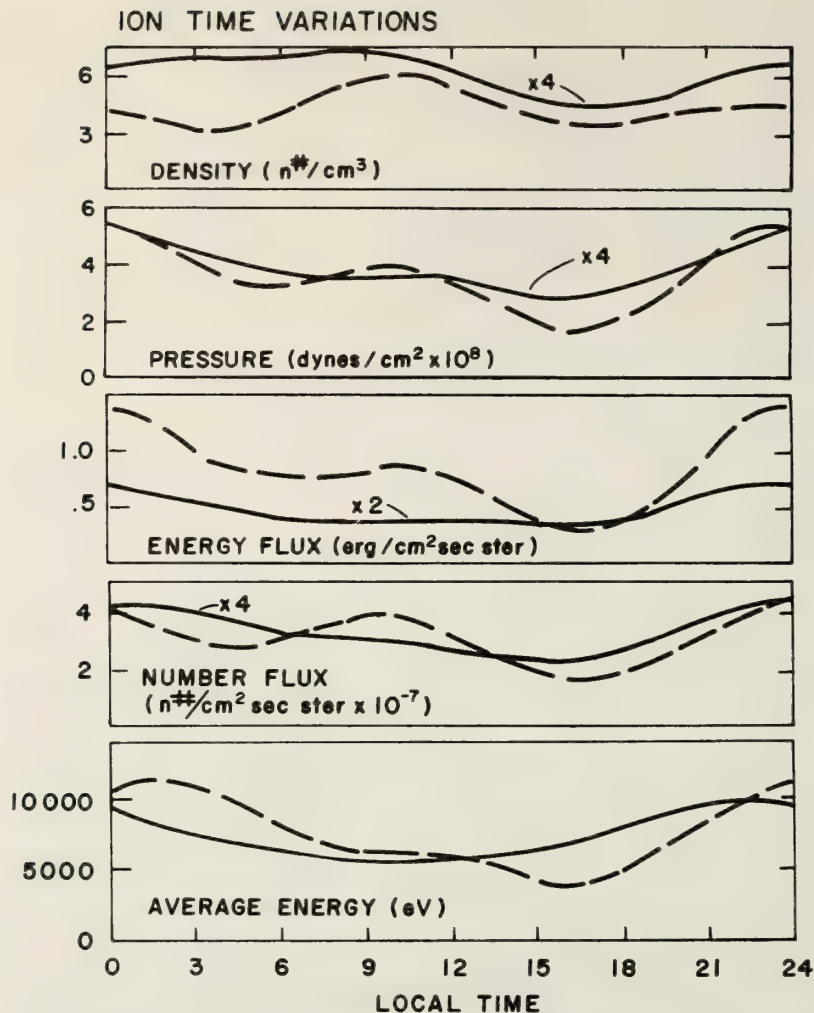


FIGURE 4b. Same as FIGURE 4a for the ions.

To use the model, one provides  $A$  and local time. The model then returns the four moments for the electrons and ion in the units given in Table 1. Typical results for  $A=15$  (average conditions) and  $A=207$  (very active conditions) are plotted in Figures 4a and b. The reader is referred to Garrett (1977) for a detailed discussion of the uses and applications of the model and problems associated with it. Only a cursory description will be given here.

The analytic model has been found to adequately simulate variations in the geosynchronous environment following substorm injection when a satellite is at local midnight for  $A$  values of  $\sim 4$  to  $\sim 48$  (i.e., low to moderately high activity). Above  $\sim 48$ , the model simulates properly the environment but is biased toward the plasma parameters on the particular days of high activity that were studied, not average conditions. The parameters returned by the model show the peak in spacecraft charging to shift from near-midnight (as expected) to near-noon for high levels of activity. This was traced to the actual data for which the plasma parameters clearly peak near noon for high activity (i.e., days 348, 1970; 217 and 223, 1972). Although we

were careful to select only days for which injections began when the ATS-5 satellite was near-midnight; this may well be a common feature of the plasma conditions associated with high activity.

### PREDICTION RESULTS

As the preceding discussion indicates, we have two different types of models of the geosynchronous environment and two different models of the spacecraft charging phenomenon. In Figure 5 we have attempted to bring together these diverse elements into a consistent picture of how the potential varies in eclipse at geosynchronous orbit as a function of the geomagnetic activity as represented by  $K_p$ . Figure 5 is somewhat complicated so we will dwell at some length on its  $P$  derivation and interpretation.

Ten minute averages of the satellite potential while in eclipse were estimated for seven eclipse seasons for ATS-5 (1969 to 1972) and one for ATS-6 (1976). These data were averaged over each eclipse period. The

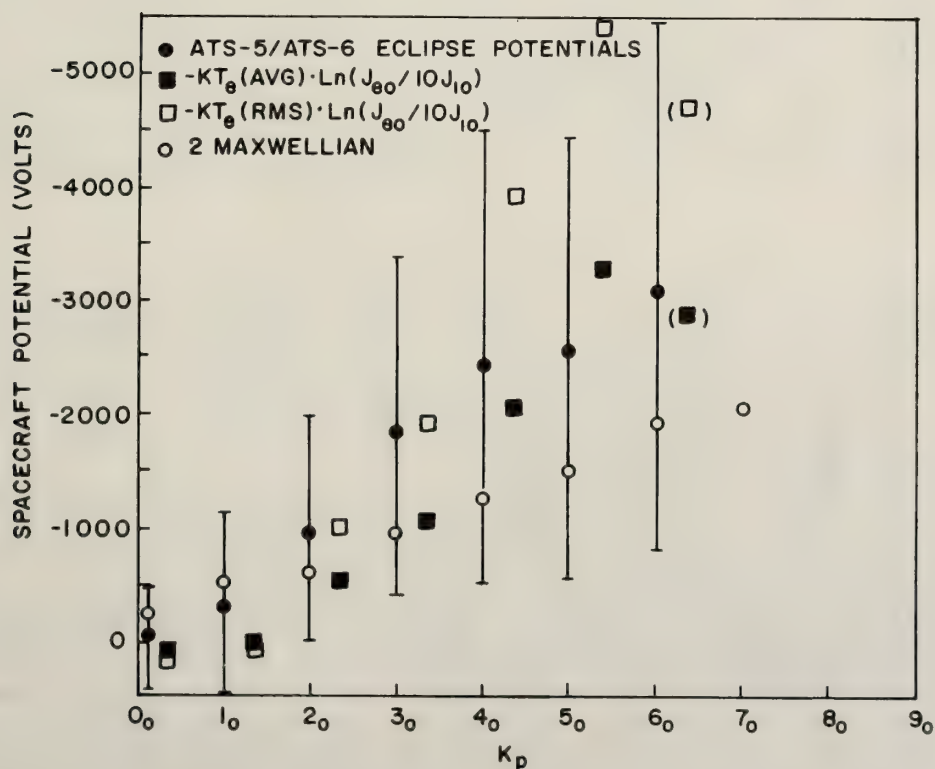


FIGURE 5. The spacecraft potential in eclipse as a function of  $K_p$  for actual observations (closed circles) and for various theoretical predictions. All plasma conditions and potential values have been included (i.e., zero potential values have been included).



results, averaged by intervals of  $K_p$  (0- to 0+, 1- to 1+, etc.), are plotted as the closed circles in Figure 5. Although large error bars exist, there is a clear linear relationship between eclipse potential and  $K_p$ . In the following we will combine the environmental and charging models  $P$  in order to test the validity of this relationship.

As represented by Figure 6, we have plotted the average values of  $T_e$  (RMS) as a function of  $K_p$  for the ATS-5 data in Figure 3. Combining these results with similar plots of  $J_e$ ,  $J_i$ , and  $T_e$  (AVG) from the same ATS-5 data base for the time interval 2100-0300 local time (i.e., the region where eclipses occur), it is a simple matter to estimate the eclipse potential according to equation 8. The results for  $T_e$  (AVG) (closed squares) and  $T_e$  (RMS) (open squares) bracket the observation's in Figure 5 - the  $T_e$  (RMS) values being somewhat higher while the  $T_e$  (AVG) values are somewhat lower. Considering the assumption's used in equation 8, we find this to be a remarkable confirmation of the simple relation between  $K_p$  and potential observed by ATS-5 and ATS-6

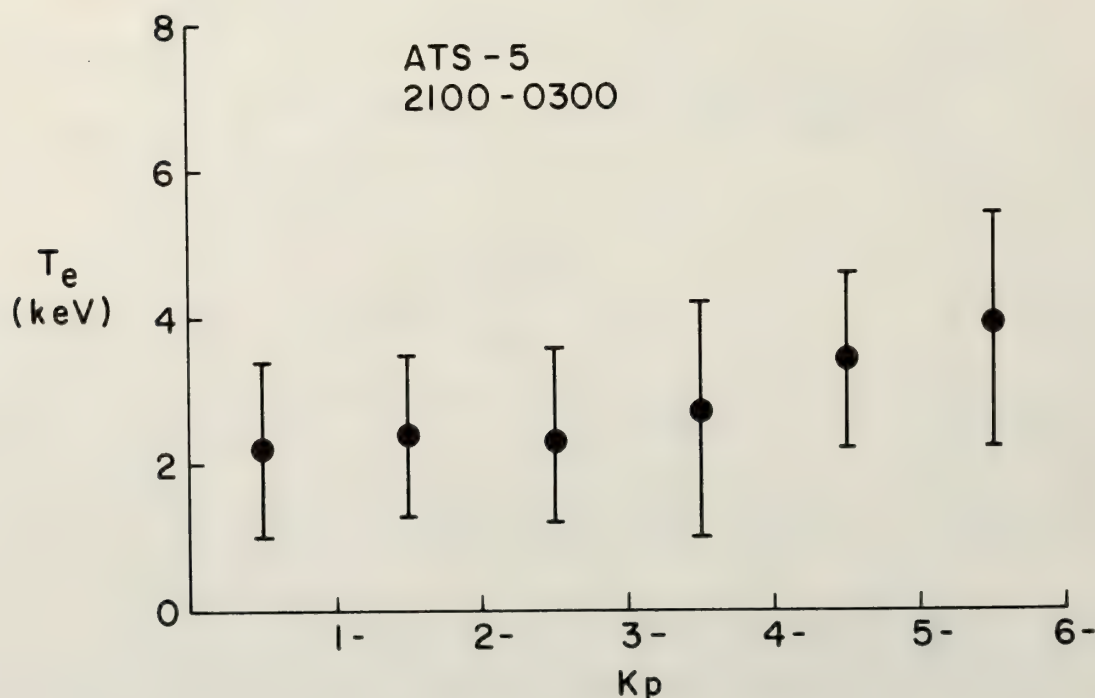


FIGURE 6. The average and standard deviation of  $T$ (RMS) for the electrons between local times at 2100 - 0300 as a function of the geomagnetic activity index  $K_p$ . The data are from 50 days of ten-minute averages of ATS-5 measurements.

We can also employ the more sophisticated models of spacecraft charging and of the environment to test these results. Assuming plasma conditions for a local time of 0130, we found the eclipse potential based on a two Maxwellian plasma distribution. The geosynchronous simulation model is only related to the statistical model in that it should, hopefully, reproduce the average plasma conditions in a form consistent with the statistical model. Likewise, the potential model is only related to equation 8 in that it is based on the same fundamental assumption of a thick sheath. These two model sets should represent significantly different ways of computing the relation between  $K_p$  and eclipse potential. The results (the open circles) in FIGURE 5 speak for themselves - even though somewhat lower than the results predicted by equation 8: they are in good agreement with the actual ATS-5 and ATS-6 observations.

### CONCLUSION

Three different methods have been presented that indicate a strong linear correlation between spacecraft potential (as measured by ATS-5 and ATS-6) in eclipse at geosynchronous orbit: by observation, by a simple plasma model combined with statistical data, and by a sophisticated charging model combined with a simulation of the geosynchronous environment. Our intent has been to not only to scientifically analyze this phenomenon, but to illustrate by example AFGL's developments in the areas of predicting spacecraft potentials and the environment. In several recent papers (Garrett, 1977; Garrett and Rubin, 1978; Garrett and DeForest, 1979), we have illustrated other capabilities of these models, such as the ability to simulate potential changes in the earth's penumbra. It is hoped that the results of this paper and those cited will prove useful to others in predicting spacecraft potentials and the low energy geosynchronous environment.

### REFERENCES

- DeForest, S.E. (1972): Spacecraft Charging at Synchronous Orbit, J. Geophys. Res., 77 (No. 4):751.
- Garrett, H.B., (1977): Modeling of the Geosynchronous Orbit Plasma Environment - Part I, AFGL-TR-77-0288.
- Garrett, H.B. (1978): Spacecraft Potential Calculations - A Model, AFGL-TR-78-0116.
- Garrett, H.B. (1979): Review of Quantitative Models of the 0 to 100 keV Near-Earth Plasma, to appear in Rev. Geophys.
- Garrett, H.B., Mullen, G., Ziemba, E., and DeForest, S.E. (1978): Modeling of the Geosynchronous Orbit Plasma Environment - Part 2, A statistical Atlas, AFGL-TR-78-0304.
- Garrett, H.B., and Rubin, A.G. (1978): Spacecraft charging at geosynchronous orbit-generalized solution for eclipse passage, Geophys. Res. Ltr., 5(No.10), 865.
- Garrett, H.B. and DeForest, S.E. (1979) Analytical Simulation of the Geosynchronous Plasma Environment, to appear in Planet. Spac. Sci.
- LaFramboise, J.G., and Prokopenko, S.M.L. (1977): Numerical simulation of spacecraft charging phenomenon, Proc. of the Spacecraft Charging Conference, AFGL-TR-77-0051/NASA TMX-73537.



- Parker, L.W. (1977): Calculation of sheath and wake structure about a pill-box-shaped spacecraft in a flowing plasma, Proc. of the Spacecraft Charging Conference, AFGL-TR-77-0051/NASA TMX-73537.
- Rothwell, P.L., Rubin, A.G., and Yates, G.K. (1977): A simulation model of time dependent plasma-spacecraft interactions, Proc. of the Spacecraft Charging Conference, AFGL-TR-77-0051/NASA TMX-73537.
- Rubin, A.G., Rothwell, P.L., and Yates, G.K. (1978): Reduction of Spacecraft charging using highly emissive surface materials, Proc. 1978 Symposium on the Effect of the Ionosphere on Space and Terrestrial Systems, Naval Res. Lab., Washington, D. C.
- Whipple, E. C. (1965): The Equilibrium Potential of a Body in the Upper Atmosphere, NASA X-615-65-296.

THE TERRESTRIAL RADIATION ENVIRONMENT AND EVA'S:  
PREDICTION REQUIREMENTS, MODEL IMPROVEMENTS, AND WARNING SYSTEMS

E. G. Stassinopoulos  
National Space Science Data Center  
NASA Goddard Space Flight Center  
Greenbelt, Maryland 20771

The advent of the "Shuttle Era" predicates increased human presence in space with an attending potential for substantial Extra-Vehicular Activities (EVA's), whose frequency and duration may well exceed anything previously experienced on APOLLO or Spacelab missions. The total medical-biological impact of the earth's space radiation environment on humans is, of course, a function of combined EVA and non-EVA exposure. In either case, the correct assessment of the eventual health risk to crew members is crucial to the success and viability of a project or mission. It may be a major, if not a critical, factor in determining feasibility. It will directly affect long range planning and crew scheduling by establishing safe crew rotation limits.

Aside from the medical-biological aspect itself, the validity of any assessment depends entirely on the existence of good and reliable models providing the high quality data that is needed for such evaluations, which should contain time histories of storm and substorm events, their intensities, their frequency of occurrence, their duration, etc. Along these lines, some thoughts are presented on prediction-requirements, on advantageous and desirable model developments and improvements, and on systems that need to be designed and tested, which would alert space crews and maintenance personnel about impending radiation danger.

The radiation hazard to a crew member is, of course, greatest during an EVA period when he has left the safety of his habitat (living-module or transport-vehicle). Then he is most vulnerable, unshielded and exposed as he is to the full impact of the "actual" radiation environment, afforded only the marginal protection of his space-suit. The emphasis on "actual" is intended to stress the difference and the contrast to an "average" environment, such as is given by most current models, where short term fluctuations and excursions (of the order of hours, days, or weeks) are masked by the averaging process.

Since in terms of magnetospheric processes and solar flares, very quiet or very perturbed monthly periods are not uncommon, and since a crew rotation period may well be of that order, it is most desirable and advantageous to have models that address such time resolutions. The availability of this type of data may considerably improve planning capabilities, allowing for flexibility in crew rotation (in response and according to absence or occurrence of events), and design of a safer and more reliable warning system. However, under present state-of-art conditions of quantitative modeling of magnetospheric processes and until more sophisticated dynamic models become available, mostly averaged data are provided by existing models. Hence, these averaged data will have to be used in current predictive studies and analyses. For reference purposes these data are designated "Base Line Values" (BLV).



Therefore, in future mission-evaluations, the safety and protection of shelters, shields, suits, the exposure duration and tolerance of humans, the mission dose limits, etc. etc., will be determined by the BLV's. Any substantial upward deviation from these BLV's must then be considered alarming and potentially harmful to the EVA crews, while variations in the quiet-time background (below the averages) may be disregarded.

Finally, in regards to an EVA warning system, three essential problems come to mind:

1. An on-board system may not provide an adequate safety margin because the actual termination of an EVA (retreat into shelter) may require a certain amount of time (to turn off instruments; to retrieve and store equipment; to shut, couple, or secure loose parts, tools, hatches etc.; to move to shelter which could be a considerable distance), by which time the event may already be upon the station.

2. An overly sensitive system is not necessarily an advantage because a response to weak (or preliminary, or unconfirmed) warning signals may result in many false alarms and many unnecessary or premature EVA terminations with a considerable loss of time and at substantial additional cost to the project. The triggering of the alarm on the basis of certain event precursors alone may be an excessive precaution: the expected full event-impact and predicted consequences may not materialize for that particular event at the specific station location.

3. The spatial and temporal dependence of a warning system may have to be determined. That is, will the measurements made by detectors at different locations and the signals received from them at the EVA location be compatible? At all times? Will the receiving EVA controller be able to relate them to the same event, within the proper time frame, predicting the same effects?

Specifically, are distance and direction of detector separation important? Will local-time dependencies affect one or both detectors? If both, then equally so? Will high latitude (and polar) phenomena distort or influence the measurements by the one but not the other detector? Or are high latitude (and polar) signals better and more reliable? Is there a time-delay or phase-lag to be considered? That is, are signals simultaneous or need they be synchronized? To this extent, it may be necessary to develop a system whereby the warning interpretation will be identical for both detectors (spatial correlation) as to whether a hazardous event is actually occurring, when it may impact the EVA location (temporal correlation), and how it may develop in time.

# DOSIMETRY IN THE MANNED SPACE PROGRAM

William Atwell  
Rockwell International  
Houston, Texas

## INTRODUCTION

The purpose of this paper is twofold. First, we present an historical review of the space radiation dosimetric results of the various U.S. manned spaceflight programs. A discussion of the instrumentation, space radiation environments, spacecraft shielding models, analytical dose calculations, and comparisons with actual measurements is included in the review. Second, an outline is presented of the future requirements of the solar-terrestrial data user who will support manned spaceflight operations during the Space Shuttle era.

One of the problems recognized early in the manned spaceflight program was that of space radiation effects on crewmen and onboard equipment, such as photographic film and radiation-sensitive experiments. Generally, space radiation includes the geomagnetically-trapped (Van Allen) proton and electron radiation, solar cosmic radiation, and galactic cosmic radiation.

The Space Shuttle era will usher-in a wider range of space missions. Long-term space exposure will be required for space construction programs, such as the Satellite Solar Power Station program and, perhaps, space colonization. In the geosynchronous environment, for example, solar particles are virtually unimpeded by the earth's magnetic field (Paulikas and Blake, 1971) and could pose a serious radiation hazard. Also, solar activity causes short-term enhancements in the trapped electron population that could have deleterious effects on extra-vehicular-activity (EVA) crewmen wearing thinly-shielded spacesuits.

## HISTORICAL REVIEW

### MERCURY PROGRAM

Early proponents of manned spaceflight were alarmed when information transmitted from the first three Explorer satellites, launched during the first half of 1958, disclosed the existence of a huge envelope of radiation beyond the ionosphere. This radiation, consisting of protons and electrons trapped in the earth's magnetic field, later became known



as the Van Allen belts, after its discoverer, Dr. James Van Allen, University of Iowa. The Mercury flights were planned low earth orbit flights that would fly "under" the Van Allen region. As a precaution passive thermoluminiscent dosimeters (TLD) were worn by the Mercury crewmen, and the radiation doses, in the millirad (mr) range, were attributed to penetrating protons.

An enhancement in the trapped electron environment caused by the high-altitude nuclear detonation on July 9, 1962, necessitated the real-time monitoring of the radiation exposure on Schirra's MA-8 flight, since the last three orbits passed through the South Atlantic Anomaly (SAA). However, measurements indicated that the astronaut received less than 30 mr body surface dose (Warren and Gill, 1964). Warren and Baker (1965) report approximately 40 mr maximum dose on the MA-9 mission.

By combining selective mission planning with the intrinsic radiation shielding characteristics of the spacecraft, it was concluded that man could perform effectively in the space environment with minimal radiation risk.

### GEMINI PROGRAM

A joint National Aeronautics and Space Administration (NASA) - Department of Defense (DOD) program was initiated to develop models of the earth's trapped radiation environment (Vette, 1965) prior to Gemini. The first model to be developed during this effort was the AE1 electron and AP1 to AP4 proton environments (Vette, 1966). A solid angle sector analysis (spacecraft shield distribution) was performed on the Gemini spacecraft by Chappell, et al. (1964). The spacecraft shield distribution and the trapped radiation model were used to determine the expected radiation doses (aluminum spherical shell shielding over  $4\pi$  solid angle) for the various Gemini missions. The calculational approach is shown in Figure 1.

The Gemini radiation monitoring system (GRMS), which consists of two 10-cc tissue-equivalent ionization chambers, was developed and flown on Gemini X and XI missions to provide real-time readout capability, since these two missions went higher into the Van Allen belt than man had ever been before. The GRMS measurements compare reasonably well with the TLD measurements. The differences are attributed to variations in local shielding by the spacecraft (Richmond, 1972). Table I lists the radiation doses measured with the passive TLD's worn by the Gemini crewmen; the GRMS measurements are noted in the table.

In 1962 the Space Science Board recommended to NASA the establishment of radiation dose guidelines for manned spaceflight exposure limits. The guidelines were updated in 1970 and are shown in Table II (Space Science Board, 1970).

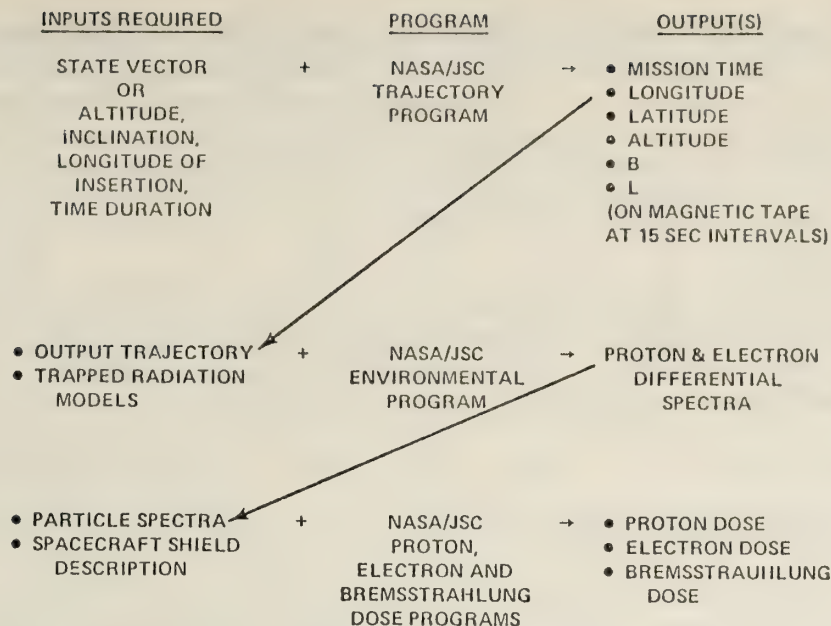


Figure 1 - TRAPPED RADIATION ANALYTICAL DOSE COMPUTATIONS

Table I - GEMINI ORBITAL PARAMETERS AND AVERAGE TLD RADIATION DOSES<sup>\*</sup>

GEMINI MISSION	LAUNCH DATE	APOGEE, N.MI.	PERIGEE, N.MI.	NUMBER OF REVOLUTIONS	INCLINATION, DEG	COMMANDER DOSE, mr	PILOT DOSE, mr
III	MAR 23, 1965	121	87	3	32.5	20	31
IV	JUNE 3, 1965	152	88	62	32.5	42	50
V	AUG 21, 1965	189	87	120	32.5	185	167
VII	DEC 4, 1965	177	87	206	29	161	166
VI	DEC 15, 1965	140 177	87 87	16	29	25	24
VIII	MAR 16, 1966	161	86	7	29	< 10	10
IX	JUNE 3, 1966	168	86	45	29	17	22
X <sup>a</sup>	JULY 18, 1966	412 161	161 86	8 35	29	685	768
XI <sup>b</sup>	SEPT 12, 1966	739 161	161 87	2 42	29	30	27
XII	NOV 11, 1966	163	87	59	29	< 20	20

<sup>\*</sup> Richmond, 1972, a - GRMS - 910 mr, b - GRMS - 30 mr



Table II - MISSION OPERATIONAL DOSE (MOD) GUIDELINES\*

MOD, REM	SKIN, 0.1 mm	EYE, 3 mm	BONE MARROW, 5 cm
30 DAY MAX.	75	37	25
QUARTERLY MAX.	105	52	35
YEARLY MAX.	225	112	75
CAREER LIMIT	1200	600	400

\* Space Science Board, 1970

### APOLLO PROGRAM

The commitment of the Apollo lunar landing program to land man on the moon and return him to earth safely necessitated in flying through the intense Van Allen region. In addition, the crew would be exposed to the full brunt of galactic cosmic radiation, and solar proton radiation-- should a solar proton event occur. Radiation dosimeters on the unmanned Apollo 4 and 6 missions indicated radiation levels in the spacecraft would produce tolerable crew doses (White and Hardy, 1969). The Apollo spacecraft is a rather thickly-shielded spacecraft and the transit through the Van Allen region is rapid (few hours). There were 11 manned Apollo missions; Apollo 7 and 9 were low-earth orbit missions primarily to check out the Crew Service Module (CSM) and Lunar Module (LM), respectively. Apollo 8 and 10 were non-lunar landing, lunar orbit missions to carry out systems checkout. Apollo 11, 12 and 14-17 were successful lunar landing missions; Apollo 13 was terminated prematurely when a tank failed in the Service Module (SM).

There were several types of radiation monitoring systems flown during Apollo:

- Nuclear Particle Detection System (NPDS) - a proton-alpha spectrometer carried in the SM that telemetered flux and spectral data.
- Van Allen Belt Dosimeter (VABD) - a skin - and depth-dose-rate instrument, fixed in the Command Module (CM) that telemetered dose rates.
- Radiation Survey Meter (RSM) - a portable, hand-held, tissue-equivalent dose-rate meter carried in the CM and transferred to the LM for lunar operations.

- Personal Radiation Dosimeter (PRD) - integrating, tissue-equivalent ionization chambers carried by each crewman; designed to measure skin dose.
- Passive Dosimeters (PD) - TLD packets worn by the crewmen.

See Richmond, et al. (1968) for a detailed description of each instrument.

Premission radiation doses were calculated at the VABD location using a 968 volume element shield description of the Apollo spacecraft (Liley, 1968). Figure 2 shows the areal density distribution at the VABD in the Apollo CM. This type of plot has obvious merit in that the thinly-shielded portions of the spacecraft can be responsible for a significant portion of the total dose.

The TLD dose results and a comparison of the computed and measured VABD radiation doses for the Apollo missions are shown in Table III. The variations in dose from mission to mission are due to trajectory differences through the Van Allen region. Again, the differences between VABD and TLD values are attributed to variations in local shielding by the spacecraft and crewmen.

The Solar Particle Alert Network (SPAN) was established to provide the Space Environment personnel in the Mission Control Center-Houston (MCC-H) with solar activity information in the form of solar optical and radio frequency data (Higgins, 1965). In the event of a major solar flare

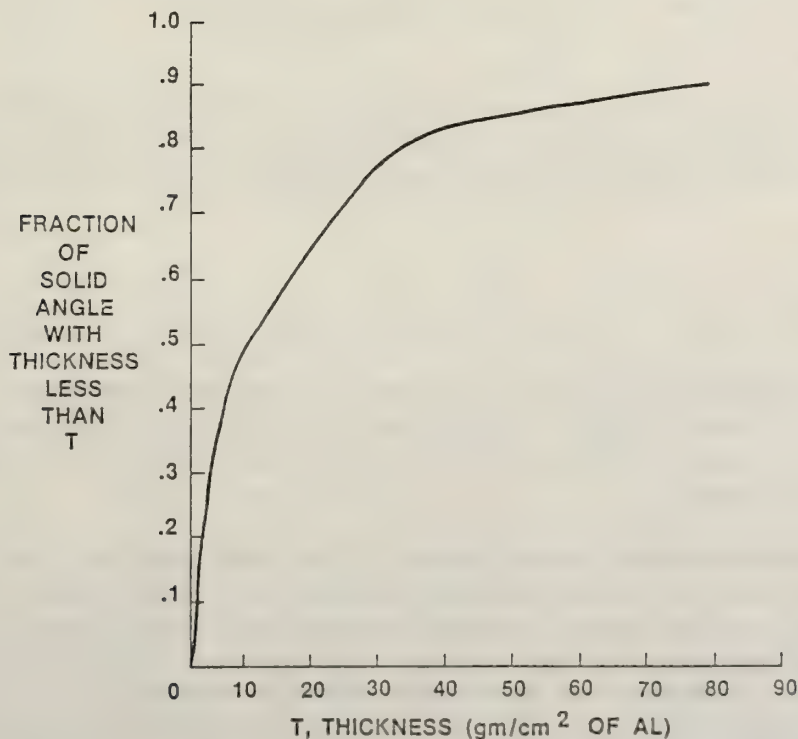


Figure 2 - AREAL DENSITY DISTRIBUTION AT THE VABD IN APOLLO CM



Table III - APOLLO TLD AND CUMULATIVE VABD DOSE COMPARISONS

Apollo Mission	TLD Skin* Dose, rad	Van Allen Belt Dosimeter (VABD)**			
		Computed		Measured	
		Sensor No. 1 <sup>a</sup>	Sensor No. 2 <sup>b</sup>	Sensor No. 1	Sensor No. 2
7	0.16	NA	NA	NA	NA
8	0.16	0.26	0.11	0.17	0.15
9	0.20	NA	NA	NA	NA
10	0.48	0.68	0.47	0.66	0.43
11	0.18	0.21	0.19	0.20	0.18
12	0.58	0.92	0.62	1.20	0.73
13	0.24	0.46	0.33	0.31 <sup>c</sup>	0.22 <sup>c</sup>
14	1.14	0.83	0.57	0.82	0.51
15	0.30	0.36	0.32	0.32	0.28
16	0.51	0.61	0.45	0.49	0.36
17	0.55	0.64	0.47	0.51	0.37

\* - English, et al, 1973; Bailey, 1976

\*\* - White, et al, 1972

a - Ion chamber shield thickness - 0.23 gm/cm<sup>2</sup> of aluminum (skin)

b - " " " " - 6.34 " " " (depth)

c - Data is incomplete due to loss of telemetry

and the confirmation of solar particles from satellite and/or NPDS data, a dose projection would be made and submitted to the Flight Surgeon. However, no major solar proton events occurred during the Apollo program.

### SKYLAB PROGRAM

The Skylab program consisted of three, three-men missions flown during May 1973-February 1974. Skylab was in a 435 km orbit at 50 degrees inclination. The groundtrack was such that the orbit became repetitive every five days.

The spacecraft shielding description of the Apollo CSM used during the Skylab program was the same as that used for the Apollo program. The Orbital Assembly (OA) shield model containing 3355 volume elements was prepared by Lockheed-Georgia under contract to the NASA/Marshall Space Flight Center. Premission dose calculations were generated at various locations in the CSM and OA using these two models with the AP1, 5, 6 and 7 proton models (Vette, 1966; King, 1967; Lavine and Vette, 1969, 1970) and the AE4 and AE5 electron models (Singley and Vette, 1972;

Teague and Vette, 1972). Agreement between calculated (premission) and measured (actual) was considered very good.

Onboard radiation instrumentation consisted of crew-worn TLD passive dosimeters, personnel radiation dosimeters (PRD), Van Allen Belt Dosimeter (VABD), radiation survey meter (RSM), and an electron-proton spectrometer (EPS). The TLD's were read out by radiological health personnel after each mission. The TLD dose results are shown in Table IV. The VABD and EPS data were telemetered to ground support personnel in near real-time, and the PRD readings were reported by the crew on a daily basis. The RSM was used for vehicle area monitoring during extra-vehicular activities.

#### APOLLO SOYUZ TEST PROGRAM

The ASTP was a joint United States-Russian space rendezvous mission. The U.S. crewmen were Stafford, Slayton and Brand; launch occurred on July 15, 1975. The mission duration was 9 days at an altitude of 225 km and 51.8° inclination. Each of the three U.S. crewmen was provided with a PRD and a TLD and these were identical to the units used on Skylab. A daily readout of the integrated dose recorded on the PRD gave the ground monitoring team a check against the calculated radiation dose projections and verified that the radiation environment had not changed from the anticipated levels. Postflight analysis of the TLD's indicated a radiation dose of approximately 100 mr for each crewman (Bailey, 1976).

Table IV - SKYLAB TLD DOSES<sup>\*</sup>

Skylab Mission No.	Launch Date	Duration, days	Crew TLD Dose, rad		
			Commander	Science Pilot	Pilot
1 <sup>+</sup>	May 14, 1973	-			
2	May 25, 1973	28	1.62	1.66	1.81
3	July 28, 1973	59	3.67	3.73	4.21
4	Nov. 16, 1973	84	8.02	7.36	6.80

\* - Bailey, 1976

+ - Launch of the Orbital Workshop



## SPACE SHUTTLE ERA

The first six Space Shuttle missions have been designated as the Orbital Flight Test (OFT). These will be low earth orbit missions ranging in altitude of 150-250 nmi at  $40.5^\circ$  inclinations; the mission duration will range from 30 hours to 7 days. The basic mission objectives will be various systems checkouts, which are essential in the initial phases of any program.

A preliminary shield model description of the Space Shuttle Orbiter, which contains 1400 volume elements, has been recently completed (Beever, 1979). Shield distributions have been generated for several locations of interest; some of these include the mission and science specialists' work stations, at the eye level of the commander and the pilot, and the six tentative locations where active, self-integrating dosimeters will be flown. The nominal daily radiation dose exposure at these locations was calculated using the AE6 (Teague, et al., 1976) and AEI7 (Teague, 1977) electron models and AP8 (Sawyer and Vette, 1976) proton model. For the OFT initial computations at these various locations indicate radiation dose rates in the "few mr/day" range.

For Space Shuttle support the Natural Environment Support Room (NESR) located in the MCC-H is being established to provide flight support personnel with solar and space environment data. In addition, data analysis, formatting, distribution, real-time event analysis and forecast services will be performed by NESR support personnel. The NESR will provide support to various functional groups such as Payload Operations Control, Flight Control, Surgeon, and Mission Operations Planning.

Basically, the data available to mission support users will consist of both meteorological and space environment data. The space environment data will be used to support two areas:

- data from conventional observations available for payload support
- data from space radiation observations for detecting hazardous conditions in support of all Shuttle flight operations

A basic set of support data will be obtained in the form of normal observations, analyses, and forecasts for Shuttle and payload operational decisions. This basic conventional space environment information--now gathered by the National Space Environment Forecast System (joint National Oceanic and Atmospheric Administration (NOAA) and the Air Weather Service (AWS)) consists of:

- Solar patrol - nearly continuous monitoring of the sun in Hydrogen-alpha, white light, and discrete radio frequencies; real-time reports of optical and radio flares, filament and prominence activity, active region behavior, and other solar phenomena; and sudden ionospheric disturbances (SID's).

- Satellite - nearly continuous monitoring of the near-earth space environment using X-ray, particle, and magnetometer sensors on GOES (geostationary), NOAA (polar orbit) and military satellites.
- Ground based geomagnetic field monitoring, auroral observations, riometer observations, and neutron counters.
- Forecasts - daily forecasts of solar and geomagnetic activity for the next 72 hours.
- Warnings - short period forecasts or reported occurrences of major solar and solar related geophysical events.

These data will be obtained at the NESR by voice reports, teletype reports, photographs, and line drawings.

A new CRT-type terminal will be installed in the NESR for acquisition of the GOES solar, interplanetary, and geophysical data in support of payload activities. The associated solar data will also be used in the NESR for analysis of major solar events and the prediction of possible proton events for detecting hazardous conditions.

The following data will be obtained at the NESR for analysis of possible radiation hazards to the Orbiter crew, photographic film, payloads, and other equipment for every Shuttle flight.

- Solar Flare reports - size, location, time, region behavior, et cetera.
- Solar Flare data - RF and X-ray background/peak fluxes, times, et cetera.
- Energetic particle reports - satellite data.

The emphasis will concentrate on accurate forecasts and warnings. It is imperative that adequate lead-time is available to permit real-time assessment of the potential hazard so that precautionary measures can be taken by the crew.

### SUMMARY

We have presented the dosimetric results of the various U.S. manned spaceflight programs and have included a brief description of the trapped radiation and spacecraft shield models, instrumentation, and a comparison of computed and measured radiation doses. It was determined that reasonably good agreement exists between the computed and measured doses, and was based on an assessment of the nominal environment. The measured crew doses were well below the mission operational dose limits established by NASA upon recommendation of the Space Science Board.



Also, it was fortuitous that solar proton emission did not impact any U.S. manned spaceflight mission.

The Natural Environment Support Concept is being established to provide flight support personnel with solar and space environment data. The solar-terrestrial data user concerned with space radiation analysis will, as in the past, need to rely on accurate models of the trapped radiation environment. He will need accurate forecasts and warnings of both trapped and solar particle enhancements with adequate lead-time to make real-time assessment of the potential hazard, to provide management with time for decision-making, and to provide the crew with time to take precautionary measures, if required.

#### REFERENCES

Bailey, J. Vernon (1976): Dosimetry During Space Missions. IEEE Trans. on Nuclear Science, Vol. NS-23, No. 4, p. 1379-1384.

Beever, E. R. (1979): Private communication.

Chappell, D. A., R. L. Kloster, and C. H. Worneke (1964): Inherent Shielding and Dose Rate Calculations for Gemini Spacecraft. McDonnell Aircraft Corporation, St. Louis, Missouri.

English, Robert A., Richard E. Benson, J. Vernon Bailey, and Charles M. Barnes (1973): Apollo Experience Report - Protection Against Radiation. NASA TN D-7080, National Aeronautics and Space Administration, Washington, D. C.

Higgins, Peter W. (1965): Operational Procedures for Apollo Dose Reduction. In Second Symposium on Protection Against Radiations in Space (edited by Arthur Reetz, Jr.), NASA SP-71, National Aeronautics and Space Administration, Washington, D. C.

King, Joseph H. (1967): Models of the Trapped Radiation Environment, Vol. Low Energy Protons. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.

Lavine, James P., and James I. Vette (1969): Models of the Trapped Radiation Environment, Volume V: Inner Belt Protons. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.

Lavine, J. P., and J. I. Vette (1970): Models of the Trapped Radiation Environment, Vol. VI: High Energy Protons. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.

Liley, B. (1968): Private communication.

Paulikas, G. A. and J. B. Blake (1971): The Particle Environment at the Synchronous Altitude. Models of the Trapped Radiation Environment, Vol. VII: Long Term Time Variations. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.

Richmond, R. G., William G. Davis, Joseph C. Lill, and Carlos S. Warren (1968): Radiation Dosimetry for Manned Space Flight. In Protection Against Space Radiation. (edited by Arthur Reetz, Jr. and Keran O'Brien), NASA SP-169 (ANS-SO-5), National Aeronautics and Space Administration, Washington, D. C.

Richmond, Robert G. (1972): Radiation Dosimetry for the Gemini Program. NASA TN D-6695, National Aeronautics and Space Administration, Washington, D. C.

Sawyer, Donald M. and James I. Vette (1976): AP-8 Trapped Proton Environment for Solar Maximum and Solar Minimum. NSSDC/WDC-A-R&S 76-06, National Aeronautics and Space Administration, National Space Science Data Center, Greenbelt, Maryland.

Singley, A. W. and J. I. Vette (1972): The AE-4 Model of the Outer Radiation Zone Electron Environment. NSSDC/WDC-A-R&S 72-06, National Aeronautics and Space Administration, National Space Science Data Center, Greenbelt, Maryland.

Space Science Board (1970): Radiation Protection Guides and Constraints for Space-Mission and Vehicle-Design Studies Involving Nuclear Systems. National Academy of Sciences - National Research Council, Washington, D. C.

Swenson, Jr., Loyd S., James M. Grimwood, and Charles C. Alexander (1966): This New Ocean A History of Project Mercury. NASA SP-4201, National Aeronautics and Space Administration, Washington, D. C.

Teague, M. J. and J. I. Vette (1972): The Inner Zone Electron Model AE-5. NSSDC/WDC-A-R&S 72-10, National Aeronautics and Space Administration, National Space Science Data Center, Greenbelt, Maryland.

Teague, Michael J., King W. Chan and James I. Vette (1976): AE 6: A Model Environment of Trapped Electrons for Solar Maximum. NSSDC/WDC-A-R&S 76-04, National Aeronautics and Space Administration, National Space Science Data Center, Greenbelt, Maryland.

Teague, Michael J. (1977): Private communication.

Vette, James I. (1965): The Updating and Dissemination of the Knowledge of Trapped Radiation - Model Environments. In Second Symposium on Protection Against Radiations in Space (edited by Arthur Reetz, Jr.), NAS SP-71, National Aeronautics and Space Administration, Washington, D. C.

Vette, James I. (1966): Models of the Trapped Radiation Environment Volume I: Inner Zone Protons and Electrons. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.

Vette, J. I., A. B. Lucero and J. A. Wright (1966): Models of the Trapped Radiation Environment. Vol. II: Inner and Outer Zone Electrons. NASA SP-3024, National Aeronautics and Space Administration, Washington, D. C.



Warren, Carlos S. and William L. Gill (1964): Radiation Dosimetry Aboard the Spacecraft of the Eighth Mercury - Atlas Mission (MA-8). NASA TN-D-1862, National Aeronautics and Space Administration, Washington, D. C.

Warren, Carlos S. and Benny R. Baker (1965): Radiation Measurements on the Ninth Mercury - Atlas Mission (MA-9). NASA TND D-2608, National Aeronautics and Space Administration, Washington, D. C.

White, Timothy T. and Alva C. Hardy (1969): Apollo 4 and 6 Radiation Analysis. AIAA Paper No. 69-17. Presented at the AIAA 7th Aerospace Sciences Meeting, New York City, New York.

White, T. T., W. Atwell and A. C. Hardy (1972): Radiation analyses for Apollo Lunar Missions. Presented at the 18th Annual Meeting of the American Nuclear Society, Las Vegas, Nevada.

## II. GEOMAGNETIC APPLICATIONS

A WORKING GROUP REPORT prepared by W. H. Campbell, Chairman, Members: V. D. Albertson, W. M. Boerner, D. H. Boteler, W. R. Goddard, J. Hruska, S. Irwin, and J. G. Kappenman

The Working Group on Geomagnetic Applications considered three topics: electric power transmission lines, long oil and gas pipelines, and magnetic mapping for geologic and navigation needs.

### 1. ELECTRIC POWER TRANSMISSION

#### 1.1 Problems

Geomagnetically-induced currents (GIC) entering an electric power system through system grounding points can be of sufficient magnitude to cause half-cycle saturation in power transformers. This half-cycle saturation can adversely affect normal operation of a power system, as well as produce over-stressing of system equipment.

System operation can be affected in the following ways: increased reactive power demands of transformers can result in unusual and severe voltage, line flow, and frequency fluctuations; the GIC and its resultant harmonics can increase the severity of switching transients, and increase the probability of AC/DC converter system misoperation in power systems that utilize both ac and dc transmission of electric power; and GIC in current transformers, as well as harmonics due to half-cycle saturation, may cause protective relay misoperation.

Equipment stresses may occur when the increased reactive power demands of the transformers, combined with the normal load demands, exceed the load capabilities of the transformers and generators. The harmonics generated by the half-cycle saturation also may overload the harmonic filters at ac/dc converter stations. Loss of expected life of transformers may occur because of localized heating (or internal hot spots) generated by the half-cycle saturation.

#### 1.2 Present Studies of Geomagnetically Induced Currents in Electric Power Systems

Because the problems depend upon the proximity to the auroral zone and extent of the power systems, studies are being conducted in several regions (Alaska, B.C., Manitoba, Minnesota and Newfoundland) to develop models of the induction process on the basis of magnetic field data and actual induced current measurements. Induced current levels, levels of harmonics of 60 Hz. (and transformer audio levels) are being monitored for studies of relay and switching problems. A computer model is being used to study the distribution of geomagnetically induced currents in a power system and to determine the changes in reactive power flow and voltage levels that are produced. Work is also being done to estimate the probability of occurrence of geomagnetic storms that would produce induced currents and subsequent operating problems,



outages, or reduction in transformer lifetime. The solutions to problems based on these probabilities are being developed in terms of engineering techniques and system control.

Although the problem is usually not present at middle-latitude locations, the hazard there may be greater. Power companies at such locations may be unprepared for the rarely occurring, great magnetic storm that extends the strong induced currents equatorward.

### 1.3 Immediate Needs

A historical review of past disturbances on utility systems due to geomagnetic storms providing deeper insight into the statistics of disturbances on electrical transmission systems should be conducted placing particular emphasis on overall geographical location and isolated regions of abnormally high disturbances. GIC effects should be analyzed relative to other simultaneous disturbances such as lightning and system failures. An analysis clarifying the disturbance mechanisms as outlined above should be carried out providing hourly, diurnal, seasonal and long-term relations. We need to expand the studies of line configuration problems and GIC effects on transformers and AC/DC convertors. It would be valuable to develop standard measurement procedures for monitoring the relevant phenomena associated with the GIC electric power system problems. We need to increase the cooperative ties between groups studying the phenomena.

### 1.4 Goals

The prime goal of the research programs is to improve the stability and security of electric power transmission systems. We hope to obtain an understanding of the underlying geophysical phenomena and processes which could lead to system outages so that engineering solutions for protective systems can be developed. It is important that the results of the research efforts be communicated to the scientific community and the affected utilities. For that purpose it is important to increase collaboration and interaction between the government, university and utility operating staff.

## 2. OIL AND GAS PIPELINES

### 2.1 Problems

The principal concern for geomagnetic effects is the degree of corrosion that may occur on high latitude oil and gas pipelines as a result of induced currents during geomagnetic storms. Although the general magnitudes of induced currents are known now, the preferential locals of current flow between the pipe and ground are still not understood; it is at these places at high latitudes that corrosion is important. At low latitudes, the induced currents can interfere with the corrosion monitoring systems. Some pipeline electrical monitoring system failures have been experienced, these occasions require improved assessment of the possible contribution of high, induced currents to such failures.

## 2.2 Present Studies

Geomagnetically-induced currents (GIC) on pipelines are now being monitored at auroral and mid-latitudes by pipeline operation companies. One research group is investigating the broad scale distribution of current along the Alaska oil pipeline and attempting to predict expected levels of maximum current surges. Another group is looking into the local current flow concentration between pipe and ground and the extent of corrosion at auroral zone sites.

## 2.3 Immediate Needs

Specific information on the localization of current flow is needed. A study of the effects of earth-surface geology, topology and climatic effects upon the induced currents is required. It would be helpful if the available pipeline operational logs were used to review the absence (or presence) of geomagnetic effects upon pipeline functions.

## 2.4 Goals

One objective of the research program is to answer questions concerning the effect and origin of induced currents upon the operation, security, and corrosion of the long oil and gas pipelines. There is a need to understand the physical and chemical processes that occur on the pipelines in such a way that adequate engineering solutions to problems may be obtained. A general goal is to improve the cooperation and communication of study results between industry and researchers.

# 3. MAGNETIC MAPPING FOR GEOLOGICAL AND NAVIGATIONAL REQUIREMENTS

## 3.1 Problems

There is a need to improve the selection of appropriate days for magnetic field mapping. Induction effects upon mapping observations need to be better separated from remnant magnetization effects. Improved methods for removal of the main field and its secular variations are required. A more complete evaluation of the limitations that natural geomagnetic noise imposes upon mapping observations is needed.

## 3.2 Present Studies

Secular field change modeling is under constant development and study. The natural field variations are being investigated from the viewpoint of stochastic process noise. Improved methods of establishing quiet baselines from secular, annual, and semiannual levels are also present research topics.

## 3.3 Immediate Needs

A detailed evaluation of the limiting effect of noise upon mapping observations is needed as a function of the sampling technique for small and large scale mapping and geophysical location. Improved detailed instructions



regarding the anticipated limitations of geomagnetic disturbances upon mapping observations need to be communicated to those presently providing the mapping observations.

### 3.4 Goals

The long-term purpose of the programs would be to be able to evaluate the accuracy of old magnetic charts, to provide improved techniques for matching charts of adjacent regions and different epoch; and to improve the final chart accuracy for geologic and navigation requirements.

## 4. PREDICTION NEEDS FOR GEOMAGNETIC APPLICATIONS

Forms of predictions issued by forecast centers for geomagnetic applications should be simple and specified for individual geomagnetic zones. Both short-term predictions (72 hrs.) and long-term (27-day to yearly) are valuable. Until protective measures are developed, short-term alerts (1-3 hrs) when magnetic storms are expected to reach intense levels are needed. It would be helpful to have an auroral oval location predicted. The users would like to see a "probability index" in the warning system (e.g. "there is a 70% probability that a magnetic storm of intensity X will occur in the next 2 hrs. in geographical region Y"). For some operations, we need predictions for quiet periods. It would be helpful if the forecast centers would integrate other available station-chain data into the preparation of their forecasts. The cooperation between various forecasting center and user organizations should be improved. The forecasting centers could provide more material for the education of users in all aspects of predictions (meanings of various index changes, etc.).

## 5. ASSOCIATION ON THE EFFECTS OF GEOMAGNETICALLY INDUCED CURRENTS

Our Working Group on Geomagnetic Applications decided to form a permanent association of those concerned with the effects of geomagnetic variations upon human enterprise. Anyone interested in active or correspondence-level membership should contact W. H. Campbell at U.S.G.S. (Mailstop 964, Box 25046, Denver, Colorado 80225, USA).

# MAGNETIC STORM EFFECTS IN ELECTRIC POWER SYSTEMS AND PREDICTION NEEDS

Vernon D. Albertson  
Department of Electrical Engineering  
123 Church St. S.E.  
University of Minnesota  
Minneapolis, MN 55455

John G. Kappenman  
Minnesota Power & Light Co.  
30 West Superior St.  
Duluth, MN 55802

Geomagnetic field fluctuations produce spurious currents in electric power systems. These currents enter and exit through system grounding points remote from each other. The fundamental period of these currents is on the order of several minutes which is quasi-dc compared to the normal 60 Hz or 50 Hz power system frequency. Nearly all of the power systems problems caused by the geomagnetically-induced-currents result from the half-cycle saturation of power transformers due to simultaneous ac and dc excitation. The effects produced in power systems are presented, current research activity is discussed, and magnetic storm prediction needs of the power industry are listed.

## 1. Introduction

The earliest documented cases of electric power system disturbances due to geomagnetic storms go back to March 24, 1940 on systems in northern and northeastern United States, and in Canada. Since that time, various power system disturbances have been recorded during specific magnetic storms, and from 1968 to 1972, a three and one-half year study of these effects was conducted in the United States and Canada during the peak of solar cycle 20. The major power system problems resulting from magnetic storm effects are increased system reactive power requirements, unusual real and reactive power flows, system voltage fluctuations, generation of undesirable harmonics, misoperation of protective relays, and internal localized heating in transformers.

The purpose of this paper is to provide a summary of the known effects and data, past and present research activity, and the prediction needs of the electric power industry.

## 2. Electric Power System Effects

### 2.1 Mechanism for Producing GIC in Power Systems

The principal mechanism for producing geomagnetically induced-currents (GIC) in electric power systems is the induced earth-surface-potential (ESP)



due to geomagnetic field fluctuations during a magnetic storm. The ESP, which can be 3 to 6 volts/km or higher during severe magnetic storms, is impressed across the grounded neutral points of three phase transformers as in figure 1. The grounded neutral points can be hundreds of kilometers apart, resulting in GIC's of 50 to 100 amperes in the neutral leads during severe storms. The GIC in each phase winding of the transformer causes half-cycle saturation of the ferromagnetic core, thereby triggering a host of adverse effects that are detailed later in this paper.

## 2.2 Geographical Areas Affected

Areas of the earth surface nearest to the paths of the auroral currents experience the highest values of ESP. Thus, power systems in more northern latitudes have both a higher frequency of occurrence of GIC and greater magnitudes. Another important factor in determining magnitude of the GIC is geology of the area. Areas of high ground resistivity, in particular areas of igneous rock geology, produce higher values of ESP and GIC. The documented power system effects of the August 4, 1972, storm, presented in a following section, show the influence of both latitude and igneous rock geology quite clearly.

## 2.3 Research Results from Solar Cycle 20

Coincident with the peak of solar cycle 20, the Edison Electric

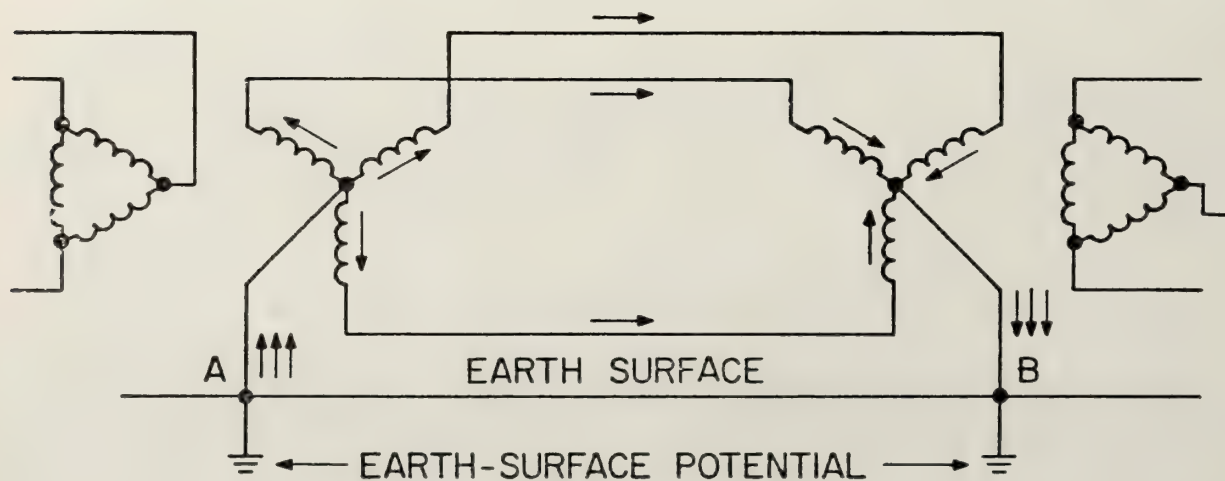


Figure 1. Mechanism for Geomagnetically-Induced-Currents to Enter an Electric Power System.

Institute sponsored a research project from 1968 to 1972 that was conducted jointly by the University of Minnesota and the General Electric Company, in cooperation with over 30 electric utility organizations in the United States and Canada. GIC recorders were installed in transformer neutral leads on power systems at locations indicated in figure 2; several recorders were also located in Canada.

A listing of the average number of GIC recordings per month, the maximum values recorded, and the mean values of the GIC, by substation, are given in table 1. Examination of table 1 reveals a dependence on latitude, but also shows that GIC are present in power systems in southern regions of the United States. In the case of the southern locations, the magnitudes of the GIC are affected by geological conditions, power network topology, and number of power system grounding points in proximity to the GIC recorder.

The most important power system effects identified as a result of this earlier research were:

- (1) Unusual real and reactive power flows.
- (2) System voltage and frequency excursions.
- (3) Misoperation of protective relays.
- (4) Transformer saturation and heating.
- (5) Generation of excessive harmonics.

#### 2.4 Effects of August, 1972, Magnetic Storm

The magnetic storm of August 4 and 5, 1972, was recorded as a K-8 storm



Figure 2. Location of GIC Recorders in the U.S. During Solar Cycle 20.



Table 1. Listing of GIC Recorder Locations During Solar Cycle 20. Sequenced by Average Number of GIC Indications Per Month from March 1969 to September 1972.

NO.	COMPANY	SUBSTATION	AVE.NO.GIC	MAX GIC	MEAN GIC
			INDIC./MO.	AMPS	AMPS
1.	Newfoundland and Labrador Power	Corner Brook	39.00	100+	16.50
2.	Minnesota Power and Light Co.	Arrowhead	6.91	58	7.65
3.	Northern States Power Co.	Red Rock	6.67	10	6.00
4.	Philadelphia Electric Co.	Peachbottom	6.35	76	10.25
5.	Philadelphia Electric Co.	Whitpain No. 3	4.95	86	9.28
6.	So. California Edison Co.	El Dorado	4.36	8	4.59
7.	Minnesota Power and Light Co.	Silver Bay	4.29	28	4.59
8.	Virginia Electric Power Co.	Elmont	4.22	68	7.11
9.	Otter Tail Power Co.	Bemidji	4.18	44	6.38
10.	Philadelphia Electric Co.	Whitpain No. 1	3.92	98	9.89
11.	Pennsylvania Power and Light	Juniata No. 2	3.83	48	8.37
12.	Metropolitan Edison Co.	Hosensack	3.76	16	5.70
13.	Pennsylvania Power and Light	Juniata No. 1	3.75	44	11.42
14.	Central Maine Power Co.	Wyman Hydro	3.35	32	7.11
15.	So. California Edison Co.	Mammoth Pool	3.33	4	4.00
16.	So. California Edison Co.	Lugo	3.27	18	8.25
17.	Otter Tail Power Co.	Winger	2.89	20	5.66
18.	Commonwealth Edison Co.	Crawford	2.53	12	4.91
19.	Arizona Public Service	Pinnacle No. 2	2.50	8	4.16
20.	Central Main Power Co.	Bucksport	2.41	32	6.73
21.	So. California Edison Co.	Sylmar	2.40	22	7.84
22.	Texas Electric Service	Everman	2.34	24	6.33
23.	Wisconsin Power and Light Co.	Port Edwards	2.31	8	4.67
24.	New England Electric System	Pratts Junction	1.89	10	4.96
25.	Consolidated Edison Co.	Pleasant Valley	1.82	68	13.40
26.	Northern States Power Co.	Black Dog	1.81	16	6.25
27.	Northern States Power Co.	Minnesota Valley	1.47	16	8.40
28.	Consumers Power Co.	Riggsville	1.34	8	6.34
29.	Utah Power and Light Co.	Naughton	1.33	12	6.67
30.	Duke Power Co.	Beckerite Tie	1.33	12	4.78
31.	Consolidated Edison Co.	Sprain Brook	1.30	32	7.78
32.	Virginia Electric Power Co.	Dooms	1.17	15	6.29
33.	Duke Power Co.	Marshall Steam	1.06	8	6.29
34.	Bonneville Power Adm.	Aberdeen	.95	12	4.80
35.	Duke Power Co.	Eno Tie	.93	12	5.89
36.	New England Electric System	Tewksbury	.91	14	5.50
37.	Detroit Edison Co.	Harbor Beach	.91	8	4.80
38.	Niagara-Mohawk Power Corp.	Rotterdam	.91	18	7.11
39.	Utah Power and Light Co.	Sage	.82	5	4.00
40.	Consolidated Edison Co.	Dunwoodie	.79	4	5.71
41.	Ohio Valley Electric Co.	Pierce B	.74	8	4.00
42.	Consumers Power Co.	Alpena	.65	12	5.60
43.	Ohio Valley Electric Co.	Pierce A	.62	8	4.21
44.	Bonneville Power Adm.	Midway	.57	10	4.50
45.	Metropolitan Edison Co.	N-Temple	.55	16	8.00
46.	Dayton Power and Light Co.	BK-S	.53	4	4.00
47.	Hydro Quebec	Boucherville No. 2	.48	8	5.33
48.	Florida Power and Light	Columbia	.46	12	5.71
49.	Pennsylvania Electric Co.	Shawville No. 4	.36	12	6.40
50.	New England Electric System	Comerford	.32	6	4.50
51.	Hydro Quebec	Boucherville No. 3	.32	8	6.00
52.	No. Indiana Public Service	Plymouth	.26	<4	<4.00
53.	Dayton Power and Light Co.	BK-N	.25	4	4.00
54.	Pennsylvania Electric Co.	East Towanda	.20	4	4.00
55.	Consolidated Edison Co.	Rainey	.19	6	4.00
56.	Arizona Public Service	Pinnacle No. 1	.19	9	6.00
57.	Texas Electric Service	Graham	.16	4	4.00
58.	Utah Power and Light Co.	Ovid	.13	<4	<4.00
59.	Pennsylvania Electric Co.	Shawville No. 1	.11	4	4.00
60.	Bonneville Power Adm.	Redmond	.07	<4	<4.00
61.	Duke Power Co.	Mitchell River Tie	.07	<4	<4.00
62.	Detroit Edison Co.	Tuscola	.07	<4	<4.00
63.	Florida Power and Light Co.	Starke	.06	<4	<4.00
64.	Texas Electric Service	Jewett	.05	<4	<4.00

at the Fredericksburg, VA, observatory. Power system disturbances were noted on a large number of power systems in the U.S. and Canada during this storm; figure 3 shows the locations of reported power system disturbances.

The effects of the storm on power systems were most severe in north central United States and Canada. Examples of specific system disturbances are listed in table 2. The phenomenon of unusual watt and var flow, voltage fluctuations, relay operation, excessive harmonics, and frequency deviations are all evident. One utility in Canada had two generating units trip out, and narrowly averted additional generator unit tripouts. No large area black-outs or system instability occurred, but the potential exists for such problems under certain system loading and operating conditions during severe magnetic storms.

### 3. Present Research Activity

#### 3.1. Introduction

This section presents a brief overview of present known research activities, in the U.S. and Canada, that are related to the effects of geomagnetic storms on electric power systems. No attempt is made to list research that may be underway outside the U.S. and Canada. Also, the authors apologize for unintentional omissions from the listing.



Figure 3. Reported Power System Disturbances on August 4, 1972.  
▲ = strong. ▲ = moderate. △ = weak. Igneous rock areas outlined.



Table 2. Specific Power System Disturbances During K-3 Storm of August 4, 1972.

Manitoba Hydro Co., Winnipeg, Manitoba, Canada

USA Tie Line (La Verendrye) 120 to 164 to 44 MW  
 USA Tie Line (La Verendrye) +25 to -100 MVAR  
 Grand Rapids 230 kV bus 295 to 336 to 295 to 340 MW  
 Grand Rapids 230 kV bus 40 to 350 MVAR  
 Grand Rapids 230 kV bus 237 to 220 to 247 kV  
 La Verendrye 110 kV bus 114 to 96 to 116 kV  
 Morden 115 kV bus 112 to 96 kV  
 Parkdale 115 kV bus 113 to 96 kV  
 Selkirk 115 kV bus 116 to 100 kV  
 System Voltage 111.0 to 90.0 kV

At La Verendrye, SIC ~ 1100 amps. in neutral of  
 230/115 kV transformer #2  
 At Grand Rapids, SIC ~ 1100 amps. in neutral of  
 13.8/230 kV transformer #1

System Frequency 60.00 to 59.95 to 60.08 Hz

Seven Sisters Unit #2 (25 MW) tripped by generator field  
 ground and overcurrent time A and C phase relays.

Slave Falls Unit #3 (9 MW) tripped by undervoltage relay.

Grand Rapids reported phase unbalanced annunciations on  
 Units #1, 2, 3, and 4.

Kettle Rapids reported phase unbalanced annunciations on  
 Units #2 and 5 (R24C line).

Grand Rapids and Kettle Units went to maximum VAR  
 output at 17:40 CDT, August 4, 1972.

Load Dispatch requested all generator units to maximum  
 voltage boost at 1755 DCT, hoping to prevent loss of  
 additional generator units.

No transformers tripped out by differential relay opera-  
 tions.

Minnesota Power and Light Co., Duluth, Minnesota

Aurora Station 30 to 87 MW  
 Clay Boswell Gen. station 25 to 74 MVAR  
 Blackberry 230 kV substation 40 to 100 to 50 MVAR  
 Blanchard 34.5 kV bus 36 to 30 kV  
 Boswell 115 kV bus 122 to 116 kV  
 Hibbing 115 kV bus 120 to 112 kV  
 Hibbard 115 kV bus 115 to 105 kV  
 Thomson 46 kV bus 46.4 to 43.2 kV  
 Pillager 34.5 kV bus 36 to 30 kV  
 Knife Falls Hydro Station 2.3 to 2.2 kV  
 Little Fall Hydro Station 2.4 to 2.03 kV

Riverton Substation 400 amperes of third-harmonic cur-  
 rent in auto-transformer tertiary

Blackberry Substation 300 amperes of third-harmonic  
 current in auto-transformer tertiary

Notice no phase current unbalance or transformer neutral  
 current

Arrowhead Substation, Duluth, Minnesota: SIC first went  
 to 22 amperes at 1555 PDT, 60 amperes at 1742 CDT  
 and dropped exponentially to 20 amperes at 1845  
 CDT, then varied from 5 to 20 amperes until 2100  
 CDT

Northern States Power Co., Minneapolis, Minnesota

King - Eau Claire 345 kV line 116 to 92 MW  
 Red Rock - Adams 345 kV line 0 to 30 MW  
 Red Rock - Adams 345 kV line 90 to 100 MVAR  
 Lakefield - Parkers Lake 345 kV line 50 to 60 MW  
 Lakefield - Parkers Lake 234 kV line 110 to 50 MVAR  
 Eau Claire - Rocky Run 345 kV line 50 to 80 MW  
 Red Rock - Arrowhead 230 kV line 50 to 70 MW  
 Red Rock - Arrowhead 230 kV line 0 to 50 MVAR  
 Riverton - Monticello 230 kV line 40 to 20 MW  
 Fargo - Cass Co. 115 kV line #1 52 to 46 MW  
 Fargo - Cass Co. 115 kV line #2 24 to 12 MW  
 Maple River - Red River 115 kV line 45 to 40 MW  
 USBR - NSP Grand Forks 69 kV 12 to 15 MW  
 Red Rock 234 kV bus 349 to 345 kV  
 Red Rock 115 kV bus 119 to 118 kV  
 Grand Forks 69 kV bus 118 to 90 V  
 Cass Co. 23 kV bus 25 to 18 kV  
 Minneapolis Frequency 60.00 to 60.02 Hz.  
 Fargo, N.D. Frequency 60.00 to 60.015 Hz.

Glenwood, Minn. had an IJD differential relay on a  
 3-2.5 MVA 69-4 kV wye-wye Bank. (3-5 MVA  
 wye-delta ground banks.) No harmonic restraints on  
 relays.

Fargo, N.D. Dispatch Center emergency generator ran 5  
 minutes.

Public Service Electric and Gas Co., Newark, New Jersey

Branchburg - Ramapo 500 kV line 90 to 120 MVAR  
 Branchburg - Ramapo 500 kV line 90 to 110 MW  
 Branchburg - Alburtis 500 kV line 300 to 280 MVAR  
 Branchburg - Alburtis 500 kV line 610 to 590 to 530  
 MW

Branchburg - Whippany 500 kV line 40 to 60 MVAR

Branchburg - Whippany 500 kV line 220 to 270 MW

Linden - Goethals 230 kV line 20 to 0 80 MVAR

Linden - Goethals 230 kV line 40 to 50 MW

Pleasant Valley - Buckingham 230 kV line 10 to 20  
 MVAR

Pleasant Valley - Buckingham 230 kV line 20 to 30 MW

New Freedom - Monroe 230 kV line 80 to 60 MVAR

Roseland - Montville 230 kV line 10 to 70 MVAR

Lawrence Seitch (connected to Branchburg) 0 to 14  
 MVAR

Trenton Switch (connected to Branchburg) 23 to 21 to  
 35 MVAR

Total generation change 2900 to 3080 MW

Total generation change 1100 to 1500 MVAR

GIC at Branchburg started at about 10 amperes at 1640  
 EDT and peaked 80 amperes at 1842 EDT dropped  
 back to 5 amperes around 1855 EDT spiked occasion-  
 ally to 10 amperes and peaked again at 20 amperes at  
 2050 EDT and then subsided rapidly.

Transformer oil gas analysis from oil samples taken  
 7-17-72 and 8-16-72 showed no significant changes

Bonneville Power Administration, Vancouver, Washington

GIC recorder showed 25 amps peak with change rates of  
 2.5 to 3 amps per second

Initial excursions of 3 amps GIC began at 1735 PDT, 7  
 amps at 1810 PDT and 25 amps at 1842 PDT. This  
 diminished almost exponentially back to 3 amps GIC at  
 2100 PDT. The major portion of the disturbance was  
 over around 1940 PDT

No change in gas analysis data from oil samples taken  
 after the geomagnetic storm

### 3.2. Hydro Quebec Research Institute

The Hydro Quebec Research Institute of Varennes, Quebec, Canada, has conducted research on various aspects of geomagnetically induced currents in power systems for approximately ten years. Their most recent effort has been directed toward theoretical analysis and experimental verification of the effects of GIC in power transformers.

### 3.3. University of Alaska

The University of Alaska is conducting research into effects of GIC on oil/gas pipelines and electric power transmission systems. Geomagnetic field fluctuations and earth surface potential measurements are being correlated with GIC measured in transformer neutral leads and with zero-sequence currents measured on the Golden Valley Power Co. system at Fairbanks, Alaska.

### 3.4. University of British Columbia

Researchers at the University of British Columbia are attempting to use micropulsation indications as a precursor to the lower frequency, larger magnitude field fluctuations that cause power system disturbances. There will be cooperation with the British Columbia Hydro organization to install recorders to measure GIC on their power system. The measured GIC will then be correlated with triaxial flux gate magnetometer and micropulsation measurements.

### 3.5. University of Illinois - Chicago Circle

Researchers are seeking support for a broad research effort which would monitor GIC at critical power grid interconnections across North America. The project would utilize direct measurements of the interplanetary plasma energy flux from the ISEE-3 satellite to develop magnetospheric substorm prediction algorithms. The research might also encompass controlled ionospheric heating experiments using harmonic radiation from unbalanced power lines as trigger mechanisms.

### 3.6. University of Manitoba

This research has focused on the prediction of the magnitudes and frequency of occurrence of GIC on the Manitoba Hydro Power system. Spectral analysis of recorded GIC waveforms and of magnetograms from a nearby observatory were used in conjunction with earlier work by Campbell that related geomagnetic field variations with the Ap index.

The University of Manitoba has also developed an in-line current transducer which allows measurement of GIC directly in a transmission line conductor.



### 3.7. University of Minnesota

The University of Minnesota, Minneapolis, Minn., is presently conducting research sponsored by Minnesota Power and Light Co., Duluth, Minn., Manitoba Hydro, Winnipeg, Manitoba, Northern States Power Co., Minneapolis, Minn., and The Electric Power Research Institute, Palo Alto, California. The purpose of the study is to investigate the effects of GIC on a future 500 kV ac transmission line that will be built from Winnipeg to Duluth to Minneapolis - St. Paul. In addition to a number of engineering studies being done at the University of Minnesota, at Minnesota Power and Light Co., and at Manitoba Hydro, GIC recorders are measuring GIC before and after construction of the 500 kV line. A triaxial flux gate magnetometer and earth surface potential measuring station has been installed northwest of Duluth, Minn., within a few miles of the new 500 kV substation. The geomagnetic field and earth surface potential data will be correlated with GIC measurements. Field tests and measurements will also be made on the power system to examine increased transformer var demands and generation of harmonics.

In the last several decades most regional electric utility companies have strived to interconnect their bulk transmission systems with neighboring utilities. The formation of these networks and power pools has proven to be an effective and economical method for a localized area or system to increase reliability of service to its customers. These bulk transmission interconnections readily facilitate the exchange of economy and emergency energy between distant utility systems. This was demonstrated during the recent coal strike (winter 1978) in the eastern United States, when much of the energy consumed in the east was supplied from generating facilities in the midwest. As discussed, interconnections have in many ways enhanced the reliability and operating flexibility of electric power systems. However, in the presence of geomagnetic disturbances, degradation of system reliability can occur due to increased magnitudes of GIC caused by numerous, as well as lengthy interconnections. The earth surface potential difference generated during disturbances can cause large differences of potential from one end of a network to another, this combined with the low impedance presented by numerous interconnections and parallel paths can result in higher magnitudes of GIC than would be expected if systems were operated isolated. The proposed 500 kV interconnection from Winnipeg, Manitoba to Duluth to Minneapolis - St. Paul is one such transmission interconnection which is expected to significantly contribute to the increased occurrences and magnitudes of GIC's on the area power system. Figure 4 shows the geographical location of the line route and interconnection points. The Dorsey substation in Winnipeg will be interconnected to the Forbes substation near Duluth by 330 miles of 500 kV line. The Forbes substation will then be interconnected 150 miles to the south to the Chisago substation near the Twin Cities. Nearly the entire route of this transmission line lies within an area of igneous rock geology, compounding an already worsened situation. Also shown is the location of the Zim test site where the triaxial flux gate and earth-surface-potential recording equipment are located.



Figure 4. Geographical Location of Winnipeg - Duluth - Twin Cities 500 kV Line.



## 4. Magnetic Storm Predictions for Electric Power Systems

### 4.1. The Nature of Power System Control Centers

Power system control centers, or dispatch centers as they are also called, have the primary responsibility for operating and controlling the entire system at all times. This entails matching system generation to system load under varying system loading conditions while maintaining system frequency and voltage profiles within close tolerances. Moreover, because of the crucial nature of a continuous supply of electricity to its customers, the power system must be operated in the most reliable manner possible in the face of system disturbances such as unexpected loss of major generating units, key transmission lines, etc.

Control centers function through a combination of automatic control, either analog, digital, or both, and human operator supervision, with actual control of system equipment and elements accomplished through a complex telemetering and communication system. The degree of human versus automatic control depends on the sophistication of the system, but in any event there is some point at which the human operator can and must make decisions and exercise control over the system.

System operators are trained to respond to various credible system disturbances, such as loss of major transmission line or substation. If the control center is aware of impending system disturbances, there are usually advance procedures that can be taken that will insure that the system is in a state of maximum reliability to withstand the disturbance. Thus, for those situations in which advance warning can be given, it is important to do so.

### 4.2. Existing Power System Alerting Procedure

Based on recommendations that resulted from the research discussed earlier during solar cycle 20, there is an alerting procedure in operation for the electric utility industry. When a magnetic storm of intensity K-5 or greater is detected, the Space Environment Services Center (SESC) of NOAA, Boulder, Colorado, notifies several electric utility organizations, including the American Electric Power (AEP) Company in Canton, Ohio. The AEP Control Center is tied in to a continent-wide communication system known as the North American Power System Interconnection Committee (NAPSIC) Time Error Correction Network (TECN). Having received a K-5 (or greater) alert message from the SESC, the AEP Control Center then sends a teletype message to all North American continent utilities through the NAPSIC TECN.

The present alerting procedure is valuable to the industry, but has two disadvantages (1) the alert is a coincident-with rather than an advance warning, and (2) it provides a blanket-type coverage alert that is not essential information to all the recipients and may be ignored eventually. The prediction needs listed in the next section would overcome these disadvantages.

#### 4.3. Prediction Needs for Electric Power Systems

There are several basic requirements that should ideally be met by a magnetic storm prediction procedure for electric power systems:

1. Sufficient advance warning, say, two hours or more.
2. Expected intensity of the magnetic storm and duration.
3. Geographical areas which will be affected by the magnetic storm.

Regarding the advance warning period, two hours would be a minimum period, and something greater than two hours would be more desirable. This advance warning period is necessary so that system operators can assess their present and expected generation, transmission line flows, and loading states, and take corrective action prior to the onset of the storm. For example, if a particular system operating condition requires a heavy tie-line flow with an adjacent system at a point of interconnection, and if the interconnection is through a long east-west line, system operators may wish to reduce the tie line flow prior to the magnetic storm to minimize system stability problems. Reducing the tie line flow could involve start-up of additional generating units or negotiation for the power purchases through other interconnections.

It would also be desirable to predict the expected intensity of the storm; a K-5 storm will produce much less severe effects than a K-8 storm. Further, the expected storm intensity will determine the extent of the preparatory corrective actions taken by the system operators. Predicting the duration of the magnetic storm is also important. System operators are accustomed to tracking severe weather in their system area - the severity of the weather, the areas affected and to be affected, the rate at which the severe weather is traveling, and the expected duration of the severe weather in system subareas. In the case of magnetic storms, similar information is needed. There is also a psychological aspect to system disturbances from magnetic storms that should be mentioned. In the case of a severe magnetic storm, a system operator will probably observe unusual fluctuations in watts, vars, and voltage at all points on the system simultaneously; this is unnerving by its nature.

Past research has shown that the effects of magnetic storms on power systems can be regional in nature, if not local. Power systems in the New England States may be affected, for example, while those in north central U.S. are not, and vice versa. To eliminate unnecessary action and concern on the part of system operators, it would be most desirable if the regions to be affected could be pinpointed to some degree.

At the present time, little is known as to how a power system will behave and react to various intensities of geomagnetic disturbances. System operators can only use their intuition in adjusting conditions to lessen the impact. It is the belief of the authors that present and future research efforts, conducted by the power industry and other interested groups, will develop a method to predict the behavior of an electric power system during a storm. This method could be used similar to the way that load-flow and stability programs are now used in the power industry to predict line loading and voltage levels. When this technology has been developed, the advance warning and intensity predictions can be used by system operators in determining the system effects and the necessary corrective actions.



## 5. Conclusions

Research and operating experience has shown that geomagnetic storms cause a number of adverse effects on electric power systems. Some of these effects have the potential of contributing to serious operating problems. Advance magnetic storm warnings of the proper type could significantly reduce the risk to power systems, and it is recommended that adequate prediction procedures and methods be implemented.

## References

- Albertson, V. D., and Slothower, J. C.: The Effects of Solar Magnetic Activity on Electric Power Systems. Journal of the Minnesota Academy of Science, Vol. 34, No. 2, 1967.
- Albertson, V. D., and Thorson, J. M., Jr.: Power System Disturbances During a K-8 Geomagnetic Storm: August 4, 1972. IEEE Transactions on Power Apparatus and Systems, Vol. PAS-93, No. 4, pp. 1025-1030, July/August 1974.
- Albertson, V. D., Miske, S. A., and Thorson, J. M., Jr.: The Effects of Geomagnetic Storms on Electrical Power Systems, ibid, pp. 1030-1044, July/August 1974.
- Albertson, V. D., Clayton, R. E., Thorson, J. M., Jr., and Tripathy, S.C.: Solar-Induced-Currents in Power Systems: Cause and Effects. IEEE Transactions on Power Apparatus and Systems, Vol. PAS-92, No. 2, pp. 471-477, March/April 1973.
- Albertson, V. D., and Van Baelen, J. A.: Electric and Magnetic Fields at the Earth's Surface Due to Auroral Currents. IEEE Transactions on Power Systems, Vol. PAS-89, No. 2, pp. 578-584, April 1970.
- McNish, A. G.: Magnetic Storms. Edison Electric Institute Bulletin, July 1940.

# THE PROBLEM OF SOLAR INDUCED CURRENTS

D. H. Boteler

Department of Geophysics and Astronomy  
The University of British Columbia  
Vancouver, B.C. Canada

This paper presents a brief review of the research on Solar Induced Currents (SIC) and suggests directions that this work should take in the future. The production of surface electric fields by geomagnetic variations is discussed with reference to magnetotelluric theory; and evidence is presented to show that the severity of SIC at Cornerbrook, Newfoundland is due to channelling of currents induced in the sea through a region adjacent to the power line. An 'isolated loop' approximation is proposed for calculating the quasi-d.c. currents produced in the power system by the surface electric fields. Identification of the processes causing the geomagnetic perturbations responsible for SIC would be aided by investigation of the local time dependence of SIC occurrence. Such knowledge could potentially be used to provide several hours warning of SIC.

## 1. INTRODUCTION

Power system disturbances have been known to occur during geomagnetic storms for nearly 30 years. The disturbances are due to quasi-d.c. currents, induced in the earth by geomagnetic field variations, flowing through transformer neutral-ground connections into the power system. Because the geomagnetic storms originate with disturbances on the sun, the quasi-d.c. currents were called Solar Induced Currents (SIC), however recently the more appropriate term Geomagnetically Induced Currents (GIC) has also been used.

The level of interest of electrical engineers in geomagnetic phenomena shows a marked correlation with the sunspot cycle! Interest was first aroused by power system disturbances during the geomagnetic storm of March 24, 1940 (McNish, 1940) and extensive SIC effects were noted during the storm of February 1958 (Slothower and Albertson, 1967) and the storm of August, 1972 (Albertson et al, 1974). By the time of the last sunspot maximum (1968-1970) a major research effort had been mounted under the sponsorship of the Edison Electric Institute, and the results of this study (Albertson et al,



1973; Albertson and Thorsen, 1974) represent the principle contribution to our knowledge of SIC.

Significant contributions have been made in parallel fields by Anderson et al (1974), who studied the effect of geomagnetic disturbances on cable communication systems, and by Campbell (1978) who analysed the induced currents in the Alaska pipeline. Campbell used the geomagnetic activity index,  $A_p$ , to determine the expected levels of induced currents in the Alaska pipeline, and this technique was adapted by Goddard and Boerner (1978) to the SIC problem. Apart from this, there has been an increased awareness of the effect of geomagnetic phenomena on man-made systems (e.g. Lanzerotti, 1978) and of the effect of power system radiation on the space environment (Helliwell et al, 1975; Hayashi et al, 1978) but little work specifically related to the problem of SIC.

The facts to date about SIC are that they fluctuate with a period of several minutes, ie. are quasi-d.c. compared to 60 Hz; and their occurrence correlates with that of geomagnetic storms. Areas of igneous rock geology give rise to higher SIC values and anomalously high SIC are experienced at a location in Newfoundland (Albertson and Thorsen, 1974). SIC are generally more severe at higher latitudes and because of this are believed to be due to the geomagnetic disturbances produced by the auroral electrojet. Possible values of the surface electric field (and consequent levels of SIC) due to the auroral electrojet were calculated by Albertson and Van Baelen (1970) and shown to be consistent with observed SIC. Albertson and Van Baelen also showed that SIC should be greater in power lines running E-W compared to those running N-S, although no observation of this effect in practice has been reported in the literature.

The effects of SIC on electric power systems and the prediction needs of the power industry have been well covered by Albertson and Thorsen (1974) and Albertson and Kappenman (1979). The problem confronting geophysicists is to explain the production of surface electric fields by geomagnetic disturbances and combine the electric field information with power system parameters to produce a quantitative understanding of SIC. This knowledge should then be coupled with improvements in predicting geomagnetic disturbances to obtain forecasting of SIC levels in any particular power system.

## 2. PRODUCTION OF SURFACE ELECTRIC FIELDS

Geomagnetic disturbances induce currents in the earth and the potential drop produced by the flow of these currents can be detected at the surface as an electric field  $E$ , also called the earth surface potential, ESP. In the simplest case the geomagnetic disturbance can be considered as a downward propagating wave of frequency  $\omega$ , incident on a homogeneous earth of conductivity  $\sigma_1$ . The relationship between the electric and magnetic fields at the surface is given by

$$\frac{E_x}{H_y} = \left( \frac{j\mu_0\omega}{\sigma_1} \right)^{\frac{1}{2}} \quad (1)$$

and the depth of penetration of the wave to  $\frac{1}{e}$  its surface value is

$$\text{skin depth, } \delta_1 = \frac{1}{\sqrt{2\pi\sigma_1\omega}} = \frac{1}{2\pi} \sqrt{\frac{T}{\sigma_1}} \quad (2)$$

It is obvious that the skin depth increases with the period of the geomagnetic variation and so, in a real earth, conductivity changes with depth will affect the frequency response of the function  $E_x/H_y$ . This was recognized by Kato and Kikuchi (1950) and Kato and Yokota (1953) who expounded the initial theory which, with further developments by Cagniard (1953), led to the technique now known as the magnetotelluric method.

The magnetotelluric method uses measurements of  $E_x/H_y$  versus frequency,  $\omega$ , to determine the conductivity structure of the earth, usually by comparison of the experimental results with results from calculations for a range of earth models. In the SIC study we are concerned with the reverse problem of (hopefully) knowing the conductivity structure below the area concerned and wishing to calculate the surface electric field  $E_x$  produced by a geomagnetic variation of magnitude  $H_y$  and frequency  $\omega$ . However, the mathematical analyses of induced current developed for magnetotelluric studies (eg. Price, 1967; Jones and Price, 1970) are still applicable to the SIC problem.

The theoretical treatment of induced currents developed by Cagniard (1953) has been extended by later workers to include effects due to the scale of the source field. For example, the formula produced by Wait (1962) for a 3 layer earth with layer thicknesses  $h_1$ ,  $h_2$  and  $\infty$ , and conductivities  $\sigma_1$ ,  $\sigma_2$  and  $\sigma_3$  is

$$\frac{E_x}{H_y} = \left( \frac{j\mu_0\omega}{\sigma_1} \right)^{1/2} (1-jB)^{-1/2} Q \left[ \left( \sigma_1 \mu_0 \omega h_1 \right)^{1/2}, \frac{\sigma_1}{\sigma_2}, \frac{\sigma_3}{\sigma_1}, \frac{h_1}{h_2} \right] \quad (3)$$

where the 1st term is the formula for a homogeneous earth of conductivity  $\sigma$   
the 2nd term includes the effect of the source size  $L$ , with  $B = \delta^2 / 2L^2$   
the 3rd term contains the layered earth effects.

Price (1962) has argued that Wait's formulation contains some simplifying assumptions, regarding the effect of the source size, that affect both  $Q$  and  $\delta$ ; and there is still some confusion as to the true effect of the source size on the distribution of induced currents.

Determining the source size is also a problem. Campbell (1978) used Wait's (1962) formulation and derived the formula  $L = 0.2T$  for the scale length, in km, of a geomagnetic variation with period  $T$  sec. This formula is based on the concept that a geomagnetic variation with a period of 24 hours has a scale equal to the circumference of the earth round the auroral zone, and that shorter period variations have a correspondingly smaller scale length. However if one considers a geomagnetic bay with  $T = 1800$  secs (30 min.), produced by the auroral electrojet, Campbell's formula gives a scale length equivalent to  $7\frac{1}{2}$  degrees of longitude, whereas the auroral electrojet is known to extend up to  $60^\circ$ , or more, in longitude. The scale length of other geomagnetic variations whose origin is attributed to the auroral electrojet (eg. Pc 5 pulsations) are likely to be similarly underestimated by Campbell's formula.

Wait's formulation was applied to the SIC problem by Goddard and Boerner (1978) who used it to compute the probable frequency spectrum of electric field variations during a geomagnetic disturbance observed in Manitoba. This electric field frequency spectrum was then used for comparison with the frequency spectrum of SIC observed during the same disturbance (fig. 1). The spectra of the electric field and SIC should be highly correlated so the discrepancy shown in fig. 1 is likely due to the calculation of the electric field. Goddard and Boerner used Campbell's formula for scale length and, as indicated above, this is an approximation and may produce a different frequency



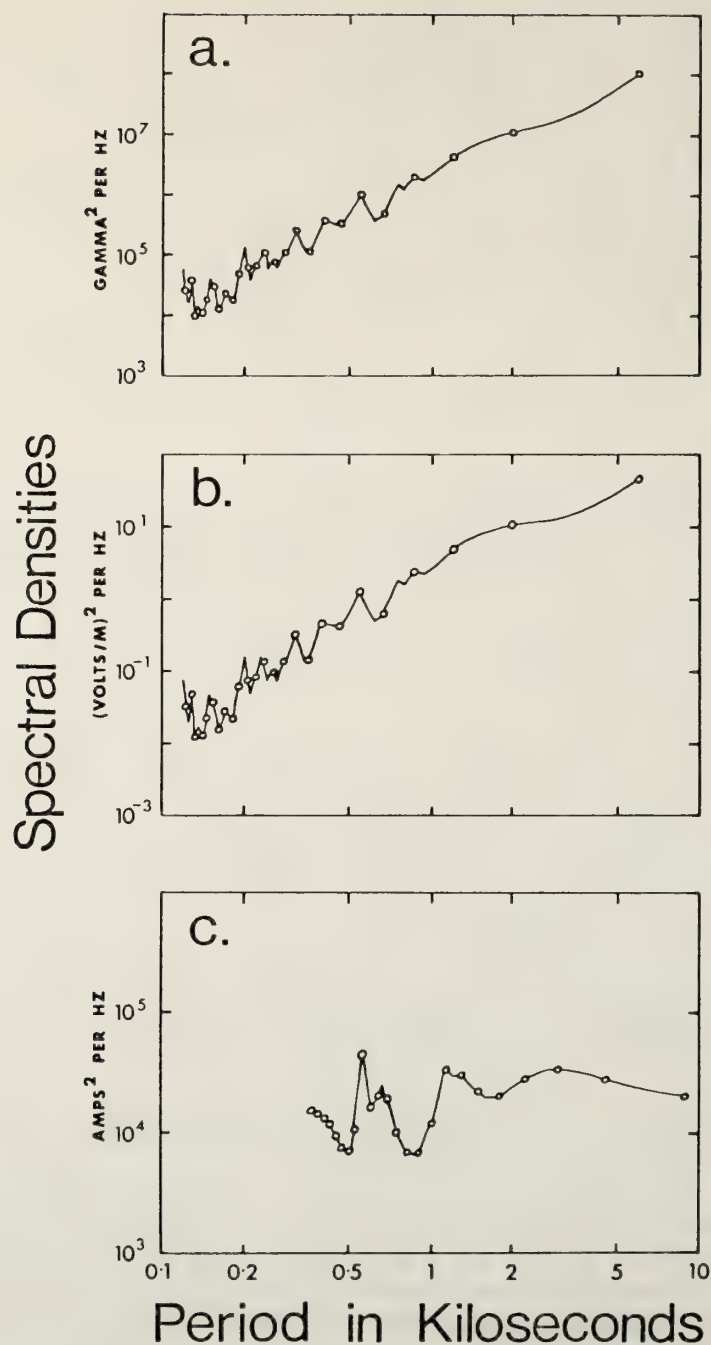


Fig. 1 Average power spectra of (a) X component of magnetic field at Whiteshell, Manitoba, (b) expected Y component of surface electric field (calculated from the magnetic field variations), and (c) solar induced currents at LaVerendrye, Manitoba, (after Goddard and Boerner, 1978).

dependence than occurs in practice. The transfer function  $E_x/H_y$  used by Goddard and Boerner, as shown by figs. 1a and 1b, is obviously  $x$  independent of frequency; however, Campbell (1978) using the same formulation, but a different earth model, obtained a transfer function dependent on frequency which he approximated by the expression

$$\left[\frac{E_x}{H_y}\right] \times 10^6 = 3.8 - 0.92 \log T \quad (4)$$

Thus the frequency dependence of the transfer function,  $E_x/H_y$ , is very dependent on the earth model used.

It should be remembered that the parameters  $\sigma_1, \sigma_2, h_1, h_2$ , etc. used in the earth models are the mean values for an area comparable to the scale of the inducing field. The significance of the conductivities at depth is in how they affect the distribution of currents with depth and consequently the value of the surface current. The electric field at a particular location can then be determined by knowledge of the surface current and the local conductivity. Igneous rock areas are an example of the effect of local conductivity because their low conductivity, compared to other rock types, gives rise to higher electric fields and in consequence a greater probability of SIC problems. In Newfoundland the SIC magnitudes are too large to be simply accounted for by low conductivity and evidence is presented in an Appendix to show that the severity of SIC at Cornerbrook, Newfoundland is due to channelling of currents induced in the sea through a region adjacent to the power line.

### 3. SYSTEM CONSIDERATIONS

The resistance of a power line, although typically a few ohms, is considerably greater than the resistance of the earth between the two ends of the power line. Thus it is applicable to calculate the surface electric field ignoring the presence of the power line, and then examine the effect of that electric field on the power line as a separate problem. The electric potential applied across the ground points at the ends of a power line is simply the product of the component of the electric field parallel to the power line and the distance spanned by the power line. However the magnitude and distribution of the currents, in the power system, produced by this earth potential depend on the relative magnitudes of the resistances of different parts of the system.

An analysis of part of the B.C. Hydro power system showed that for quasi-d.c. currents the system could be represented by a network of 'line resistances' and 'station resistances' as shown in fig. 2a. The station resistance is the resistance between the high voltage bus to which the power lines are connected and the ground mat of the substation. This station resistance could be simply the resistances of two transformers in parallel or, where autotransformers are used, a more complicated network of resistances as shown in fig. 2b. There is also a resistance between the station ground mat and 'true earth', which is called the 'station ground resistance', but this is usually small compared to the station resistance and can be neglected. The 3-phase power system can then be considered as 3 identical resistance networks in parallel and it is sufficient to consider one network only. This will enable calculations to be made of the quasi-d.c. currents expected in any part of the system; except that the current through the neutral-ground connector of a transformer (where



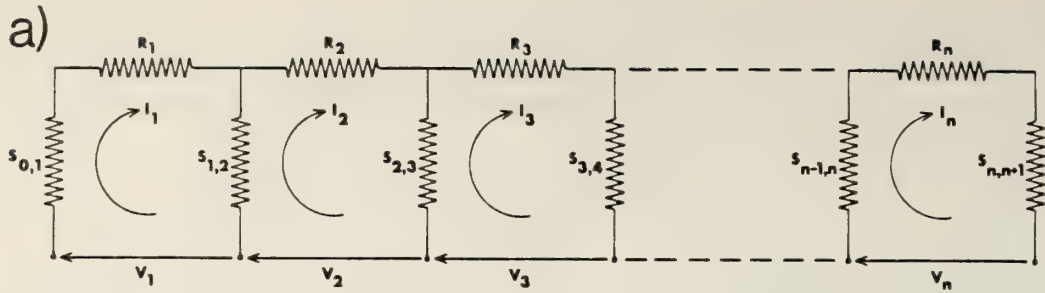
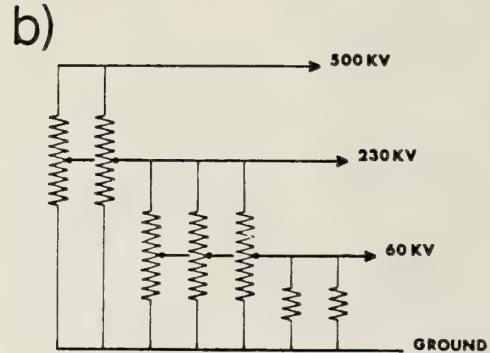


Fig. 2 a) 1-phase diagram of a power line, illustrating the network of line resistances  $R_1, R_2, \dots, R_n$ , and station resistances  $S_{0,1}, S_{1,2}, \dots, S_{n,n+1}$ .

b) Diagram of the components of the station resistance between a 500KV line and ground. (in this case at Williston substation in B.C.)



SIC are usually measured) will have currents from each of the 3 parallel networks and so will have 3 times the current calculated for an individual winding of that transformer.

To determine the quasi-d.c. currents in the network shown in fig. 2a, equate voltages round loop  $i$  where  $i=1,2,3 \dots n$ .

$$S_{i-1,i} (I_i - I_{i-1}) + R_i I_i - S_{i,i+1} (I_{i+1} - I_i) = V_i \quad (5)$$

Collecting terms gives

$$-S_{i-1,i} I_{i-1} + (R_i + S_{i-1,i} + S_{i,i+1}) I_i - S_{i,i+1} I_{i+1} = V_i \quad (6)$$

For an isolated loop

$$I_i^{(0)} = \frac{V_i}{R_i + S_{i-1,i} + S_{i,i+1}} \quad (7)$$

and this can be taken as a first approximation to the solution of the general case, ie. equation (6). Employing an iterative method one can write

$$I_i = I_i^{(0)} + I_i^{(1)} + \dots \quad (8)$$

where the first order correction  $I_i^{(1)}$  is given by

$$I_i^{(1)} = \frac{S_{i-1,i} I_{i-1}^{(0)} + S_{i,i+1} I_{i+1}^{(0)}}{R_i + S_{i-1,i} + S_{i,i+1}} \quad (9)$$

$$= \frac{1}{R_i + S_{i-1,i} + S_{i,i+1}} \left\{ \frac{S_{i-1,i}}{R_{i-1} + S_{i-2,i-1} + S_{i-1,i}} V_{i-1} + \frac{S_{i,i+1}}{R_{i+1} + S_{i,i+1} + S_{i+1,i+2}} V_{i+1} \right\} \quad (10)$$

Therefore

$$I_i = \frac{1}{R_i + S_{i-1,i} + S_{i,i+1}} \left\{ V_i + \frac{S_{i-1,i}}{R_{i-1} + S_{i-2,i-1} + S_{i-1,i}} V_{i-1} + \frac{S_{i,i+1}}{R_{i+1} + S_{i,i+1} + S_{i+1,i+2}} V_{i+1} \right\} \quad (11)$$

Of the three terms in brackets in equation (11) the coefficients of  $V_{i-1}$  and  $V_{i+1}$  will be less than one and in the many cases in which line resistances are greater than the station resistances the coefficients will be small so that these terms can be ignored. The isolated loop approximation (equation 7) then represents a reasonable solution to the problem of calculating loop currents.

To determine the current through a particular station resistance  $S_{i,i+1}$  it is necessary to consider the currents in loops  $i$  and  $i+1$ . It will be seen from equation (5) that the currents  $I_i$  and  $I_{i+1}$  tend to cancel each other at their common station, and so higher currents will be experienced in the lines than at the stations. The station resistance is actually comprised of a network such as in fig. 2b and the station current will naturally divide up between the different paths according to the relative resistances of the different transformers. Hence the current through an individual transformer will be a fraction of the current  $I_{i+1} - I_i$  flowing to ground through the station. Fig. 2b also shows that where autotransformers are used higher currents will flow in the high voltage part of the winding than in the low voltage part.

#### 4. PREDICTION OF SIC

If forecasting of SIC is to be improved it is no longer sufficient to describe SIC as being correlated with geomagnetic storms. It is desirable to ascertain which mechanisms are responsible for SIC and then prediction of these mechanisms can be related to prediction of SIC. The increase of SIC magnitudes with latitude, as mentioned earlier, points to the auroral electrojet as the cause of the field variations responsible and this has been assumed to be the case by most authors. However it is debatable whether the SIC seen at lower latitudes (eg. in the southern USA) are due to the auroral electrojet; and Andersen et al (1974) have presented evidence to show that magnetopause currents were responsible for the geomagnetic disturbance that so badly affected some communications systems in August 1972.

Resolution of this question would be greatly aided by knowledge of the diurnal occurrence pattern of SIC. Many geomagnetic phenomena have distinct average variations of occurrence with local time which are well documented (e.g. Hartz and Brice, 1967). An equivalent representation is a local time perturbation profile such as that shown in fig. 1 of Clauer and McPherron (1979). Their profile indicates that, for the current system specific (which involves a partial ring current), the maximum perturbation at mid-latitudes will occur in the afternoon. At auroral latitudes the major source of disturbances is the westward electrojet and this is located in the region extending from 22.00 to 06.00 local time. Other disturbances can be similarly associated with a range of local times and the local times of occurrence of



SIC should match those of the phenomena responsible. Also knowledge of SIC occurrence patterns at a number of locations could show whether or not the responsible phenomena change with latitude.

Should a SIC mechanism be identified which has a well-defined local time dependence this knowledge would greatly aid SIC prediction. Even if SIC are produced by each and every type of geomagnetic disturbance they may show a local-time occurrence pattern simply because of the greater frequency of occurrence or greater magnitude of one specific mechanism. In the majority of cases such a local-time dependent probability of SIC occurrence could be used to provide several hours warning to electric utility operators of potential SIC problems.

Suppose, for example, that SIC are predominantly due to geomagnetic disturbances caused by the westward auroral electrojet and so occur mainly between 22.00 and 06.00 local time. By continuously monitoring this time zone, warning of geomagnetic disturbances, in many cases, could be given before they had a significant effect in N. America. This method, of course, provides no warnings of disturbances that commence while N. America is in the 22.00-06.00 local time zone. Major disturbances will last for a day or more so the time zone specified could be considered as a "disturbed" zone,



Fig. 3 View of the earth from above the N. pole showing N. America being brought into a 'disturbed' zone by the earth's rotation.

fixed w.r.t. the sun (fig. 3), through which N. America passes as the Earth rotates. Continuous observation of this "disturbed" zone effectively requires world-wide real-time monitoring of the surface geomagnetic field and relaying of the information to a centre in N.America. (A major undertaking). However notification of disturbed conditions at European observations could be used to provide forecasts, 4 hours or more in advance, of potential SIC conditions in N. America.

Even when geomagnetic disturbances can be predicted, sections 2 and 3 have shown that there are several factors limiting the translation of this to a forecast of the SIC magnitude to be expected at any particular location. Fortunately earlier experimental work has provided us with an empirical measure of the relative magnitudes of SIC at different locations in the form of Albertson and Thorsen's (1974) cumulative severity index (CSI). SIC predictions could be broadcast for a station with a CSI of 1 and, providing a suitable scale could be devised for the predictions, the electrical utility operators at each location could multiply this by their own CSI to obtain a prediction relevant to their own system.

## 5. CONCLUSIONS

It has been demonstrated that the magnitude of SIC at different locations depends on local conditions and this is most marked in the case of the Cornerbrook power system in Newfoundland. At a particular location, the level of SIC in individual transformers (which is the critical parameter for SIC effects) depends on the relative magnitude of the currents in the adjacent parts of the power system as well as on the number and size of the transformers at the substation. The determination of the local time dependence of SIC occurrence would help identify the phenomena responsible and could potentially enable several hours warning to be given of SIC conditions.

## ACKNOWLEDGMENTS

The author would like to thank Dr. T. Watanabe of U.B.C. and Mr. R.M. Shier of B.C. Hydro for their advice and assistance. This work was supported by a Graduate Research and Engineering Technology Award sponsored by the Ministry of Education, B.C. Government and by B.C. Hydro, and by Strategic Grant G0031 of the Natural Sciences and Engineering Research Council of Canada.



## APPENDIX: THE CASE OF NEWFOUNDLAND

The severity of SIC at Cornerbrook, Newfoundland is too great to be simply another example of an igneous rock area and the case requires closer inspection. Fisher (1970) documented some of the problems due to SIC experienced with the Cornerbrook power system and he also reported that other electric utilities in Newfoundland did not experience noticeable SIC effects. The Cornerbrook power line is coincident with a zone joining carboniferous sediments to the salt water of the Gulf of St Lawrence (Wright, 1978); and Wright has suggested that this zone may be a conductive channel for telluric currents and be responsible for the great severity of SIC.

The intensity of induced currents in the oceans will be much greater than those on land because the conductivity of sea water ( $3-4 \text{ ohm}^{-1} \text{ m}^{-1}$ ) is several orders of magnitude greater than that of rock ( $10^{-1}-10^{-4} \text{ ohm}^{-1} \text{ m}^{-1}$ ). Price (1967) has shown that significant induced currents with a period of 15 min. can be expected in depths down to 500m.; therefore, for shorter periods, sizeable currents can be expected in the depths of 300m. typical of the continental shelf along the east coast of N. America. A theoretical analysis of induced currents near a conductivity discontinuity (such as a coastline) by Jones and Price (1970) shows that the currents should be concentrated on the high conductivity side of the discontinuity; and evidence for such an effect has been presented by Schmucker (1964) and Boteler (1978). Thus the indications are that high induced current densities occur along the east coast of N. America during geomagnetic disturbances.

Newfoundland and Nova Scotia represent low conductivity anomalies in the path of the coastal induced current (fig. 4.) and although the currents will tend to flow around the land the current across Newfoundland and Nova Scotia will be greater than the currents experienced on the continent proper. The currents will obviously concentrate across the narrowest parts of Newfoundland (see insets in fig 4) and the surface electric field in these locations will thus be considerably greater than elsewhere on the island. Over distances for which the potential drop in the sea is small the coastlines on opposite sides of the island can be considered as equipotentials. Hence any power line running across the island and grounded on each shore will have the same potential applied across it during geomagnetic storms. Should such power lines present the same resistance to SIC the magnitudes of SIC would be the same in each power line. However shorter lines have lower resistances and consequently higher SIC levels. Thus the Cornerbrook power line is in one of the worst possible situations for experiencing SIC problems. Another location presumably similarly affected is the Amherst area of Nova Scotia: a fact that may be significant considering the plans for power generation in the Bay of Fundy.

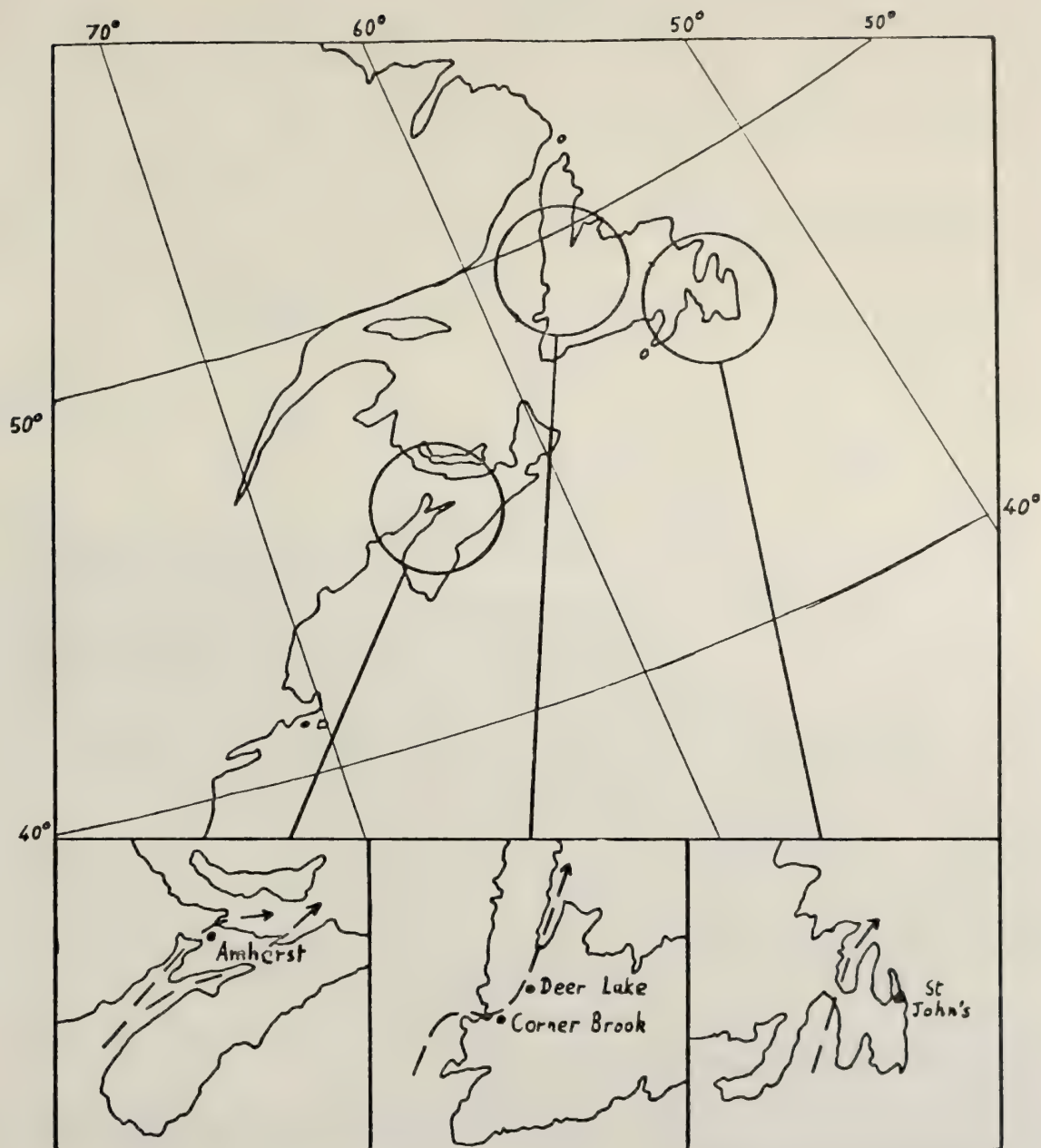


Fig. 4 Map of the E coast of N. America (geographic co-ordinates)  
The insets show the channelling of induced currents across Newfoundland and Nova Scotia that is postulated to occur during geomagnetic disturbances.



## REFERENCES

- Albertson, V.D., and J.A. Van Baelen (1970), "Electric and Magnetic Fields at the Earth's Surface due to Auroral Currents", IEEE Transactions on Power Systems, Vol. PAS-89, No. 2, p. 578-584.
- Albertson, V.D., and J.M. Thorson (1974), "Power system disturbances during a K-B geomagnetic storm: Aug. 4, 1972", IEEE Trans. on Power Apparatus, Vol. PAS-93, p. 1025-1030.
- Albertson, V.D., J.M. Thorson, and S.A. Miske (1974), "The effects of geomagnetic storms on electrical power systems", IEEE Trans. on Power Apparatus, Vol. PAS-93, p. 1031-1044.
- Albertson, V.D., J.M. Thorson, R.E. Clayton and S.C. Tripathy, (1973) "Solar Induced Currents in Power Systems: Cause and Effects", IEEE Trans. on Power Apparatus, Vol. PAS-92, p. 471-477.
- Albertson, V.D. and J.G. Kappenman (1979) "Magnetic Storm Effects in Electric Power Systems and Prediction Needs", Proc. ISTEP Workshop Program.
- Allen, J.H. (1975), "Spatial and temporal distributions of magnetic effects of auroral electrojets as derived from AE indices" J. of Geophys. Res. Vol. 80, No. 25, p. 3667-3677.
- Anderson, C.W., L.J. Lanzerotti and C.G. MacLennan (1974), "Outage of the L4 system and geomagnetic disturbances of 4 August 1972", Bell System Tech. J. 53, 1817-1837.
- Boteler, D.H. (1978), "The effect of induced currents in the sea on magnetic bays at a coastal observatory", J. Atmos. Terr. Phys. Vol. 40, p. 577-580.
- Campbell, W.H. (1978), "Induction of auroral zone electric currents within the Alaska pipeline", Pure and Applied Geophysics, Vol. 116, 31p.
- Clauer, C.R., and R.L. McPherron (1979) "Predicting Partial Ring Current Development", Proc. ISTEP Workshop Program.
- Fisher, E.D. (1970), "Influence of Solar Activity on Power Systems", Canadian Electrical Association, Nov. 1970.
- Goddard, W.R. and W.M. Boerner (1978) Report on Study of Solar Induced Currents for the Proposed 500KV Transmission Line", Univ. Manitoba Tech Report AEMSS-78.05.01.
- Hartz, T.R. and N.M. Brice, (1967), "The general pattern of auroral particle precipitation" Planet. Space. Sci., 15, No. 2, 301-329.
- Hayashi, K. et al. (1978), "Power Harmonic Radiation enhancement during the sudden commencement of a magnetic storm" Nature 275, No. 5681, p.627-629.

- Helliwell, R.A., J.P. Kalsufakis, T.F. Bell and R. Raghuram (1975), "VLF line radiation in the earth's magnetosphere and its association with power system radiation" J. Geophys. Res. 80, p. 4249-58.
- Kato, Y. and T. Kikuchi (1950) "On the phase difference of earth current induced by the changes of the earth's magnetic field (Part 1)", Science Rep. Tohoku Univ. Ser 5., Geophysics 2, p. 139-141.
- Kato, Y. and K. Yokota (1953), "Corrected paper on the phase difference of earth current induced by the changes of the earth's magnetic field", Sci. Rep. Tohoku Univ., Ser 5, Geophysics 5, p. 41-43.
- Lanzerotti, L.J., (1978), "Geomagnetic Influences on Man-Made Systems", Presented at S.T.P. Conference, Innsbruck, Austria, June 1978.
- McNish, A.G. (1940), "Magnetic Storms" Edison Electric Institute Bulletin.
- Price, A.T. (1967), "Electromagnetic Induction within the Earth", Physics of Geomagnetic Phenomena. ed Matsushita and Campbell, p. 235, Academic Press, N.Y. U.S.A.
- Schmucker, U. (1964), "Anomalies of geomagnetic variations in the south western United States", J. Geomagn. Geoelect. 15, p. 193.
- Slothower, J.C. and Albertson, V.D., (1967), "The Effects of Solar Magnetic Activity on Electric Systems", presented to the T and D Committee of EEI.
- Wright, J.A. (1978) Memorial Univ. Newfoundland. Private Communication.



# PREDICTION OF SOLAR INDUCED CURRENTS AND EFFECTS ON POWER TRANSMISSION SYSTEMS IN CENTRAL CANADA

W.R. Goddard

Department of Electrical Engineering, University of Manitoba  
Winnipeg, Canada, R3T 2N2

W-M. Boerner

Department of Information Engineering, University of Illinois  
at Chicago Circle, Chicago, Illinois 60680 U.S.A.

The extensive use of the hydro-electric generating capacity in Northern Manitoba requires long HV-DC/AC transmission lines to serve the southern part of Manitoba and neighbouring U.S. power companies. The auroral-electrojet zone covers three-quarters of the province and consequently, solar storms strongly affect these transmission lines. Harmonics are generated at transformers due to the saturation of their cores by induced currents, and the level of harmonics produced may cause malfunction of control relays, and yield unacceptable distortions in normal AC waveforms. The objective of our study is to determine the expected effects on long AC transmission systems, and in particular, a 500 KV line to be built from Winnipeg to Minneapolis-St. Paul. We have used spectral analysis of induced current records from Manitoba Hydro's LaVerendrye station and magnetograms from IMS stations in Manitoba. These analyses and results of Campbell's work on the Alaskan pipeline induction problem (1978) were used for prediction of periodic and surge currents. We conclude that the surge currents will produce significant levels of harmonics and corresponding operating problems during magnetic storms.

## 1. INTRODUCTION

In Manitoba, Canada, electric power is transmitted over vast distances from the hydro-electric generating plants in the northern part of the province primarily to serve the population in the south. The Manitoba power system has interconnections with American power networks and those of adjacent provinces in order to maintain hydro storage and take advantage of daily rate variations in thermal-electric power. For this purpose, Manitoba Hydro and Minnesota Power and Light will be constructing a 500 kilovolt alternating current between Winnipeg and Minneapolis-St. Paul. During the design phase these power authorities need to know the expected magnitude, frequency of occurrence, and duration of magnetic storms and substorms that would produce significant currents in this line. Research groups at University of Manitoba and University of Illinois, lead by Prof. W-M. Boerner, and the research group of Prof. V.D. Albertson at University of Minnesota are studying this

problem along with System Planning engineers of Manitoba Hydro. These latter engineers are concerned with ensuring the reliability of performance with minimum cost.

During magnetic storms most problems in the operation of AC power systems result from effects of transformer core saturation produced by induced currents and this saturation causes local heating and generation of harmonics. In particular undesired relay operation in transformer differential schemes is most likely, which is seen as a fault and the transformer taken out of service. Overcurrent ground relays have caused misoperation and resulted in increased zero sequence current. Cumulative heating effects on transformer insulation may reduce the lifetime of the transformer or unusual gassing during a storm may cause the transformer to be taken out of service unnecessarily. The telluric currents enter power transmission lines through the grounded neutral of wye-configuration transformers. The system we are studying has an autotransformer, and induced currents in the power network can bypass the ground connection. A 550 km length of the planned route is at an angle of about  $60^\circ$  from geomagnetic north and the total resistance of the line will be about 5.5 ohms, including terminal resistance to ground.

Our purpose in this paper is to attempt to predict the magnitudes and frequency of occurrence of induced currents in this transmission line. We have used previous works on prediction of induced currents, and our limited analyses of induced currents along with the corresponding magnetic field variations to develop methods for the desired predictions. We use a probabilistic approach to prediction.

## 2. DEVELOPMENT OF PREDICTION METHODS

A study of our prediction problem and the methods used by other researchers on related problems is necessary to establish the way in which previous results can be used and where there is a need for different techniques. Since our problem is to predict the occurrence of induced currents of sufficient magnitude to cause effects on power systems under various operating conditions, then the time and magnitude of a storm must be considered. The time of occurrence of significant currents is important since a power system may be most susceptible under certain conditions of interconnections in the network and the loading pattern. In our work on power systems we have used the simple criterion that storms should produce induced currents in excess of 20 A, since for many transformers this current produces a level of harmonic currents that may interfere with system control. Larger currents can produce instabilities that lead to generator shut-down. Events of this nature occurred during the 4 August 1972 major magnetic storm, where peak induced currents of 100 amperes were recorded by Manitoba Hydro. In Manitoba, the occurrence of substorm activity can be expected to span several hours in the interval of 3UT to 15UT. Sudden commencements may not be as significant in occurrence and magnitude, but they can occur at times that will affect the power system when loading is significant. An example of a sudden commencement effect on a power line in Manitoba has been studied by Hayashi et al (1978).

The expected power spectra of magnetic storms and substorms can be related to the planetary amplitude index and the period of the induced currents.



Campbell(1978) has obtained these empirical power spectra relationships for the Auroral zone ( $60^{\circ}$  -  $70^{\circ}$  geomagnetic N) magnetic field variations.

In addition to studying the magnetic field, from which we must determine the electric field for current induction, it is most useful to study available records of induced currents. We have obtained Manitoba Hydro's records of events for the LaVerendrye transformer neutral-to-ground connection. The observed currents are correlated with the occurrence of substorms. Thus if we have sufficient information, a relationship between the magnetic activity index  $A_k$ , magnetic field power spectra, and induced current spectra can be obtained<sup>p</sup> to predict harmonic currents on the basis of expected  $A_k$  values. We have used the power spectral analyses of three substorms and <sup>p</sup> Campbell's empirical relationships to predict the harmonic currents in the 500 KV power transmission line.

The prediction of the occurrence of very major storms and the largest electric field associated with such storms is also of interest. Campbell has also studied magnetic activity for a range of  $A_k$  index values to obtain peak-to-peak field changes as a function of the index. Siscoe(1976) has derived probability functions for the magnitudes of the three largest geomagnetic storms per solar cycle from the aa index values over nine solar cycles. This measure of storm magnitude does not indicate the rate of change of the magnetic fields. Siscoe uses the double exponential law for rare events to obtain his results. Fleming and Keller (1972) used earth surface potential measurement data to obtain the probability of large earth current storms and their results were also based on the double exponential law. However, the duration of large potentials is not studied explicitly. Since our problem requires a model for probability of both magnitude and duration of electric fields producing induced currents, we have initiated analyses of substorm magnetic variations to estimate the joint probability of the field rates of change and duration of each rate. We expect that it may be possible to use the results of the aforementioned probability models to extend our predictions to several solar cycles.

### 3. ANALYSIS FOR PREDICTION OF HARMONIC CURRENTS

We have chosen to start with Campbell's results and determine how well they suit our problem. It was expected that the analysis of substorm magnetograms from the closest magnetometer station (Whiteshell,  $59.9^{\circ}$  N geomagnetic) would give a good estimate of the desired empirical formula corresponding to Campbell's formula. His formula for the horizontal field power spectral density  $T$  (min) and magnetic activity index  $A_p$  for College, Alaska is

$$\log G(\omega) = 2.0 + 1.4 \log T + 0.043 A_p \pm 0.6 \quad (1)$$

where 1.4 is called the spectral slope since the spectra are fairly linear on a log-log plot.

As a first attempt to determine the spectral slope for Whiteshell we studied the most active three hours in substorms on 21 September 1977, 4 January 1978, and 15 February 1978. We considered both X and Y components of the magnetic field using magnetograms with one minute samples, which were provided by the Geomagnetism Division of Energy, Mines and Resources. The



individual spectra are shown together in Figure 1(a) and their average in Figure 1(b). The individual spectra tend to decrease beyond  $T=3600$  s.

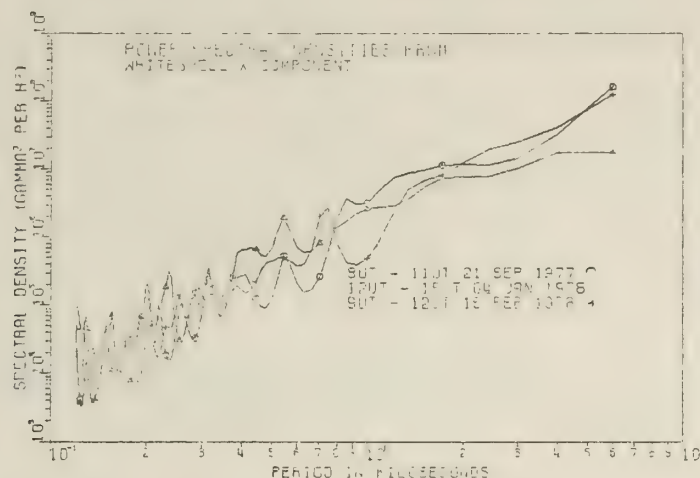


FIGURE 1(a)  
INDIVIDUAL POWER  
SPECTRA FOR THREE  
SUBSTORMS.

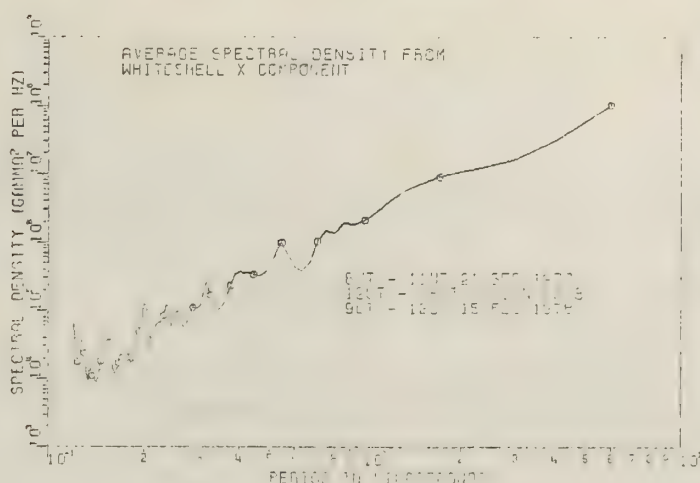


FIGURE 1(b)  
AVERAGE OF POWER  
SPECTRA FOR THREE  
SUBSTORMS.

The average spectral slope is 2.34. This slope is very close to the average slope of 2.4 obtained by Campbell and this agreement tends to support the use of his empirical relationships for our prediction problem. We were also interested in the spatial extent of the storms and we obtained data from the Fort Churchill IMS magnetometer chain from Dr. J.K. Walker of the Geomagnetism Division for Island Lake and Thompson. Figure 2 shows the individual with much lower slopes up to 1 ks and greater power in the Thompson field, particularly at shorter periods. These spectra indicate that the Whiteshell does experience variations approaching those at stations well within the auroral zone.

We would like to obtain an empirical formula for the spectral density based on Campbell's. We have obtained one estimate of the slope at similar  $a_p$  levels and since our spectral analysis was not over the entire day the  $a_p$  effective  $A_p$  index was assumed to be  $\bar{a}_p = 120$ . To give the same power spectral density as shown in Figure 1(b) at  $a_p$  the midpoint the response of the estimated relationship is

$$\log G(\omega) = 2.0 + 2.34 \log T + 0.36 A_p \pm 0.6 \quad (2)$$

where we have assumed the same constant for the tolerance.

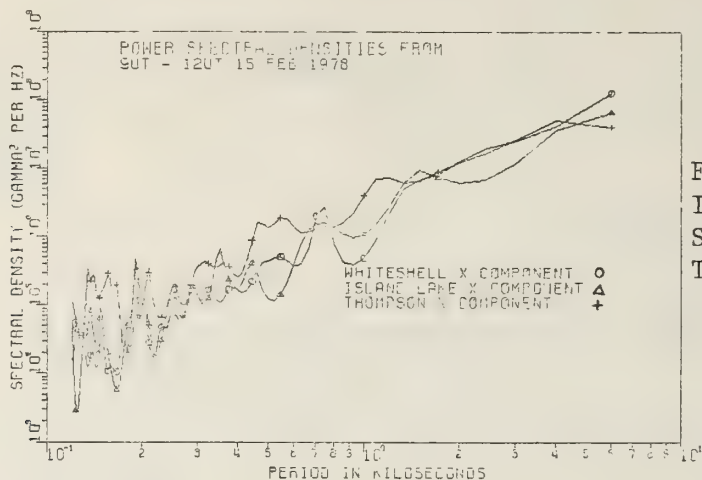


FIGURE 2.  
INDIVIDUAL POWER  
SPECTRA FOR  
THREE LOCATIONS.

The peak values of the corresponding periodic magnetic field components can be derived from the power spectrum and the X component given by an empirical relationship

$$B_x = a_T + n_T A_p \pm s_T (\text{gamma}). \quad (3)$$

We have not analyzed sufficient data to estimate these coefficients and for our results we chose to use Campbell's linear approximations to  $a_T$  and  $n_T$ , which are

$$\log a_T = 0.58 \log T - 1.33 \quad \text{and} \quad \log n_T = 0.97 \log T - 3.56 \quad (4)$$

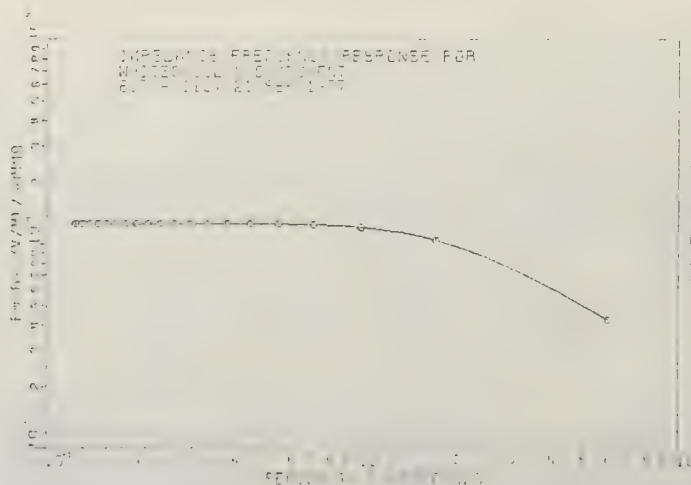


FIGURE 3.  
IMPEDANCE FOR  
OBTAINING  $E_y$   
FROM  $B_x$ .

We used a source scale of  $L = 0.2 T \text{ km}$  for  $T$  in seconds for the period of the source variation, and Wait's (1962) magnetotelluric formulation to estimate the electric field power spectra. The three layered earth model most suitable for the region of the 500 KV transmission line is  $\sigma_1 = 0.002 \text{ S/m}$ ,  $h_1 = 10 \text{ km}$ ;  $\sigma_2 = 10^{-4} \text{ S/m}$ ,  $h_2 = 390 \text{ km}$ ;  $\sigma_3 = 0.5 \text{ S/m}$ ,  $h_3 = \infty$ , corresponding to the first three layers used by Albertson (1970), where the  $\sigma$ 's are conductivities and  $h$ 's are thicknesses of each layer. The ratio of  $E/B$  for this layered earth model is shown in Figure 3. Figures 4(a) and 4(b) show the electric field power spectra for the Y and X components of the field corresponding to the X and Y components of the magnetic field at Whiteshell. The plots show upon comparison with the magnetic field spectra that the response is suppressed for periods beyond 1 ks. For 4 ks the electric field peak is 0.047 V/km and at  $T=0.4 \text{ ks}$  the peak is 0.0077 V/km.

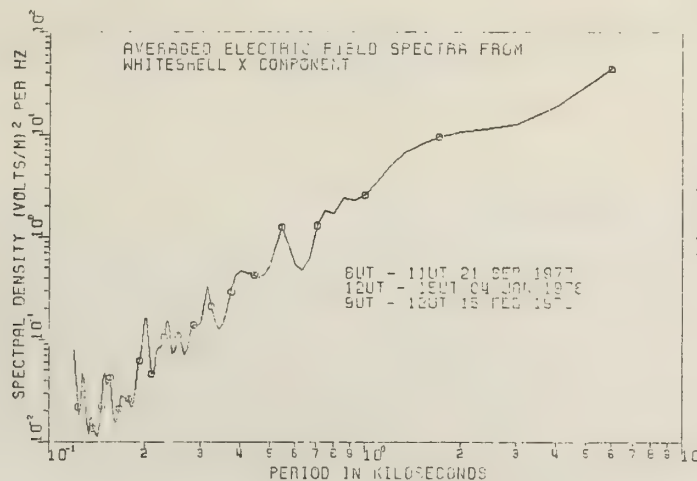


FIGURE 4(a)  
AVERAGE Y  
COMPONENT  
FROM X FIELD.

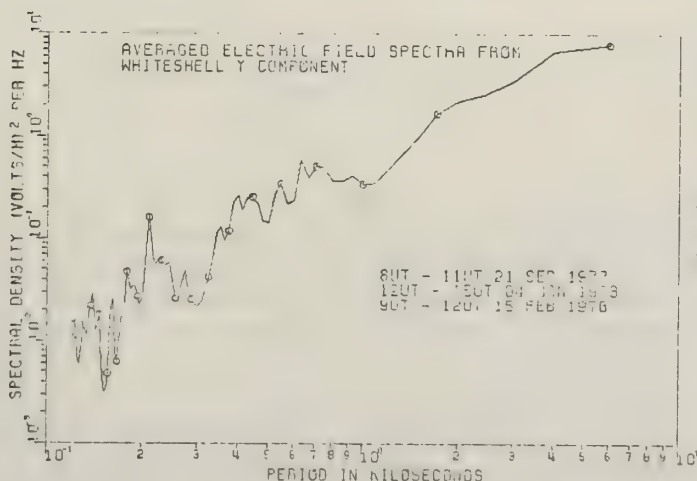


FIGURE 4(b)  
AVERAGE X  
COMPONENT  
FROM Y FIELD.

The original plan for this project included an in-line measurement device for induced currents in a 230 KV (R50L) line along a similar route. This device will be used in the near future, but since it was not available we have attempted to use recordings of transformer neutral-to-ground currents at LaVerendrye station. The individual power spectra for the three substorms were computed from 3 minute samples of analog data and these are shown in



Figure 5(a). These spectra are very different from the spectra of Figure 4,

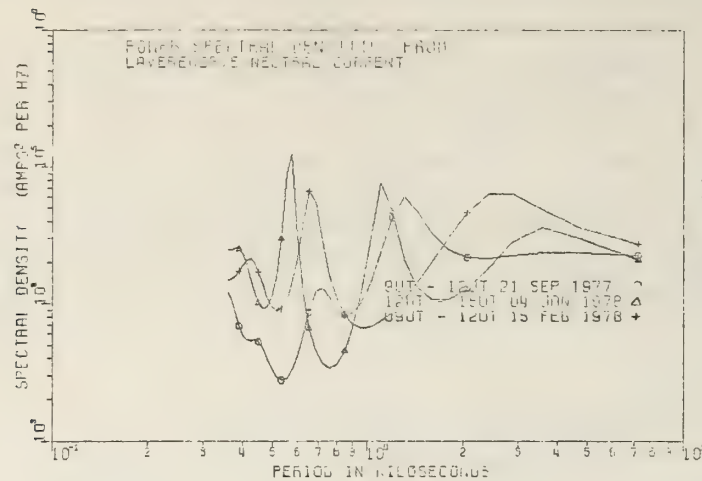


FIGURE 5(a)  
INDIVIDUAL POWER  
SPECTRA FOR THREE  
SUBSTORMS

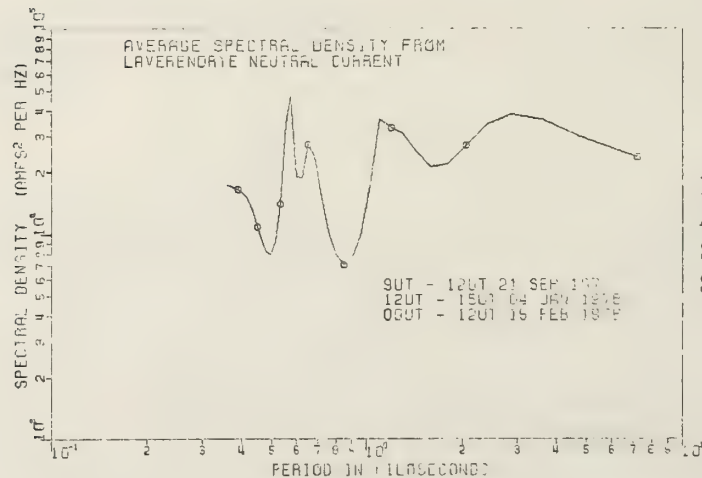


FIGURE 5(b)  
AVERAGE OF POWER  
SPECTRA FOR THREE  
SUBSTORMS.

with a few marked peaks for periods shorter than 2 ks. The trend of the spectra is towards a flat response for longer periods as shown by the average of the individual spectra in Figure 5(b). The values of the electric field with period 4 ks correspond to a Fourier component of current of 2.35 A, and a resistance per unit length of transmission line of 0.02 ohms/km, which is the correct order of magnitude for lower voltage lines. The main difficulty with the neutral current is that the contributions to the current from the system cannot be resolved into components. The approach we have used is the forward solution and perhaps we will have to develop inverse techniques for the long lines since the magnetic fields vary spatially and they are in aperiodic fields instead of the periodic ones assumed for magnetotelluric inverse modelling. The time series should be correlated to ensure that timing errors in the analog data are not occurring. In spite of these problems, we have chosen for the purposes of obtaining results in the next section to use the simple magnetotelluric model previously described.

## 4. RESULTS

### 4.1 Induction of Harmonic Currents

Using the relationship between  $B_x$  and  $A_p$  as a function of period  $T$  derived by Campbell, we have determined the electric fields for oscillating magnetic fields with periods  $T = 300s$  to  $15000s$ . We used the source scale and three layered earth model previously described. The expected currents in the 500 KV line are shown in Figure 6. This figure shows that for a line impedance of 5.5 ohms and a length of 550 km the currents range up to 11.5 A for  $A_p = 200$ , from 1.5 A when  $A_p = 20$ .

The currents during any time interval of a storm are composed of several frequencies or harmonics. The actual mean power obtained from all the harmonics is the integral or sum over the band of frequencies.

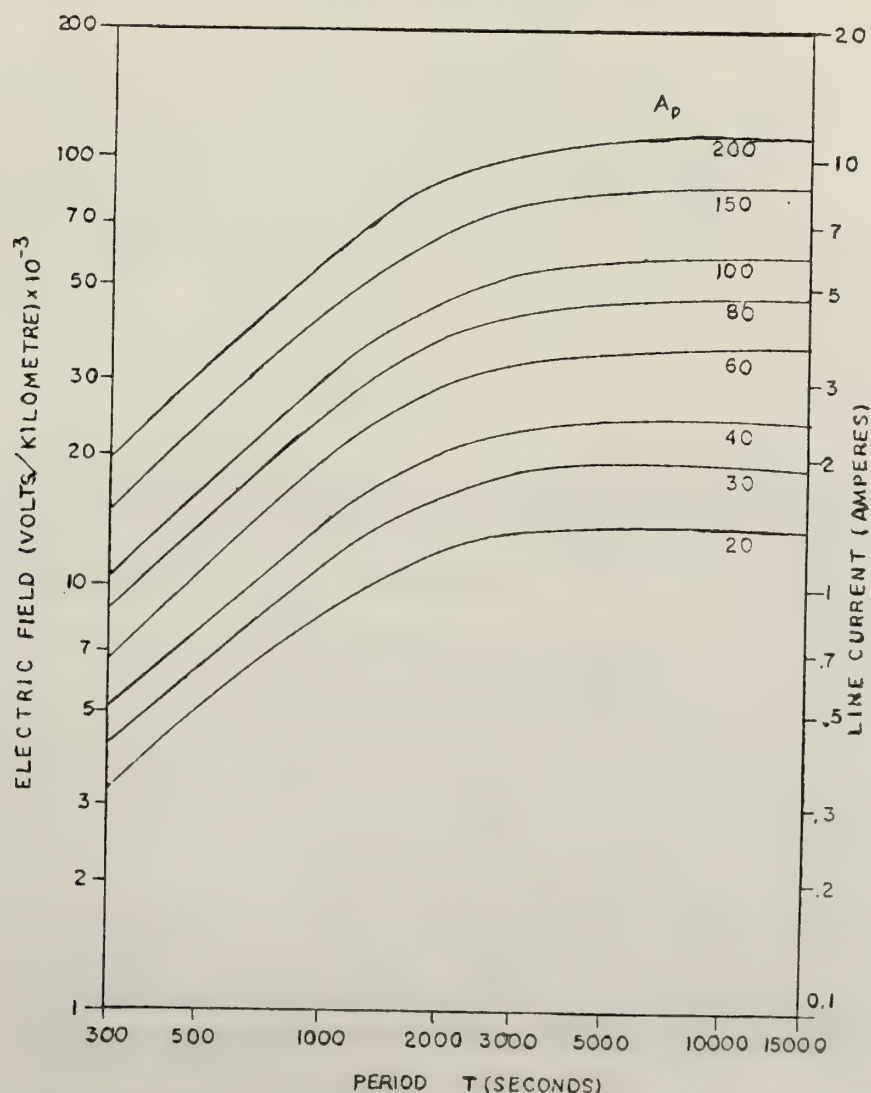


FIGURE 6. PREDICTED ELECTRIC FIELD AS A FUNCTION OF PERIOD  $T$  AND  $A_p$  INDEX, AND CORRESPONDING TRANSMISSION LINE CURRENTS.

## 4.2 Surge Currents

During each storm there are usually one or two rapid changes in the magnetic field which produce surges in the induced currents. The power spectrum obscures the existence of these surges, however, these surges could produce effects dependent on triggering mechanisms. In order to characterize these surges we have used a simple method for estimating the magnetic field rate of change and the duration  $D$  of the same rate  $R$  within a fixed root-mean-square deviation of  $10 \gamma/\text{minute}$  over the interval under test. A histogram of these rates and durations for the 21 September 1977 substorm is given in Figure 7. This result is only useful to indicate that a large

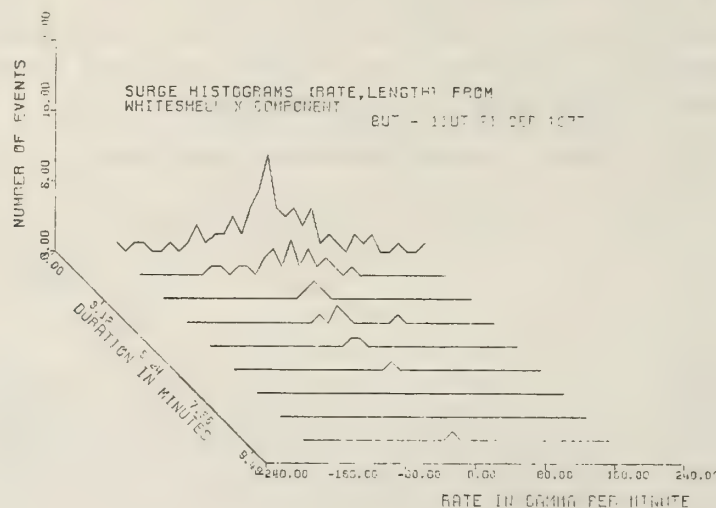


FIGURE 7. HISTOGRAM OF MAGNETIC FIELD RATES OF CHANGE AND DURATIONS FROM A MAGNETIC SUBSTORM.

ensemble of histograms may yield an estimate of a joint probability density function of the form

$$P(R,D) \propto e^{-R^2/2\sigma^2} e^{-\lambda D} \quad (5)$$

where  $\sigma$  and  $\lambda$  may be dependent on  $D$  and  $R$ , respectively. This analysis of a magnetogram indicates that one rate of change of  $+80 \gamma/\text{min}$  persisted for 4 minutes for a total change of  $320 \gamma$ . The total variation in the field was  $1400 \gamma$  during the three hour interval. One rate of  $-96 \gamma/\text{min}$  of 10 minute duration was apparently broken up into shorter intervals by the rms fit criterion used.

Clearly, we need to know the rate of change to determine the induced current, however, a model of the total variations as a function of the local  $A_k$  index due to Campbell (1978) can be used as the peak value in the harmonic models for determining the electric field. The empirical relationship is

$$B'_x = -13.4 + 14.3 A_k, \quad (6)$$

where  $A_k$  is the local equivalent amplitude index. Near the Whiteshell the



equivalent amplitude derived from the  $a_k$  index by multiplying the standard  $a_k$  for mid-latitude stations by 2.5. According to this factor the  $A_k$  could range up to 500. If rapid variations occurred, then a surge of 360 A peak would occur for  $A_k = 200$ . At longer periods for the same  $A_k$  the peak current would drop off rapidly with T, and these types of events are more likely than ones with short periods in the order of 300 seconds.

#### 4.3 Frequency of Storms

The occurrence of  $A \geq 52$  is once per 2 months on the average and a high value of  $A \geq 150^P$  is once per 2 years. Formulas are also available for days per year with a given activity level (Campbell(1978)).

#### 4.4 Expected Effects

If the cumulative effects due to the harmonic currents are shown to be significant then the active values of  $A \geq 52$  once per two months would determine the effect on the expected lifetime of a transformer.

A study performed by Manitoba Hydro (internal report) showed that harmonics are generated at unacceptable power levels when a voltage of 6 V/km is applied in a simulation of the 500 KV line. Our results at this stage show that surges with continuous currents approaching the values used in the simulation will occur at least on the average of once per 2 years.

#### REFERENCES

- Albertson, V.D., and Van Baelen (1970): Electric and magnetic fields at the earth's surface due to auroral currents. IEEE Trans. on Power Apparatus and Systems, PAS-89.
- Campbell, W.H. (1978): Induction of auroral zone electric currents within the Alaska pipeline. Pure and Applied Geophysics, Vol. 116.
- Fleming, D.B., and G.V. Keller (1972): Research on the probability of occurrence of unusually large earth-current storms. Final report Contract 14-03-79750, Bonneville Power Administration, and U.S. Bureau of Reclamation.
- Hayashi, K., T. Oguti, T. Watanabe, K. Tsuruda, S. Kokubun, and R.E. Horita (1978): Power harmonic radiation enhancement during the sudden commencement of a magnetic storm. Nature, 275:5681.
- Siscoe, G.L. (1976): On the statistics of the largest geomagnetic storms per solar cycle. J. Geophys. Research, Vol. 81.
- Wait, J.R. (1962): Theory of magneto-telluric fields. J. of Research of National Bur. of Stand. (Radio Prop.), Vol. 66D.

# HARMONICS OF 60 HZ IN POWER SYSTEMS CAUSED BY GEOMAGNETIC DISTURBANCES

K. Hayashi,\* T. Oguti,\* and T. Watanabe  
Department of Geophysics and Astronomy  
University of British Columbia  
Vancouver, British Columbia V6T 1W5

K. Tsuruda† and S. Kokubun  
Geophysics Research Laboratory  
University of Tokyo  
Tokyo, Japan

R. E. Horita  
Department of Physics  
University of Victoria  
Victoria, British Columbia

Through simultaneous VLF/ULF observations carried out near Winnipeg, Manitoba, evidence has been obtained that geomagnetic disturbances control the behavior of harmonics of 60 Hz man-made electric power. The harmonics of 60 Hz detected by the VLF receiver are at multiples of 180 Hz, indicating that they originated from a 3-phase a.c. power system. Under geomagnetically quiet conditions, only odd harmonics of 60 Hz were detected; the 3rd (180 Hz) and 9th (540 Hz) were predominant. In disturbed conditions, both odd and even harmonics were excited. Harmonics up to the 18th (1080 Hz) were detected at one time. The strength of each harmonic was found to change concurrently with geomagnetic pulsation (ULF) activity. These findings seem to indicate that a portion of telluric currents shunted into the power line system through the neutrals of the Y-connected transformers give rise to a "d.c." bias to the transformer core materials and that it distorts their hysteresis loops, activating harmonics of 60 Hz power. A mathematical proof is given that a hysteresis loop having a point of symmetry generates odd harmonics only, whereas loops lacking in point-symmetry generally give rise to both odd and even harmonics. A general formula has been obtained to calculate the strength of each harmonic based on the shape of the hysteresis loop.

---

\*On leave from Geophysics Research Laboratory, University of Tokyo  
†Institute of Space and Aeronautical Science, University of Tokyo

## 1. INTRODUCTION

The main purpose of this paper is to present several new findings about the influence of geomagnetic disturbances on power transmission systems. These findings were obtained through simultaneous VLF/ULF observations carried out by the authors in September 1977. Some of the results of the observations have been reported in Nature (Hayashi et al., 1978). This paper summarizes those results that are directly related to man-made power systems. Somewhat more elaborate discussions will also be presented about the interpretation of the observational results.

## 2. OBSERVATION

The VLF observation was performed at a temporary station set up on the shore of Lake Winnipeg, east of Riverton, Manitoba, approximately 130 km north of Winnipeg. The observational results reported herewith were obtained using a loop antenna. The ULF observation was made at three stations, all in Manitoba: Riverton; Star Lake, about 200 km east of Winnipeg; and Thompson, right in the middle of the auroral zone. A three-component induction magnetometer was set up at each of the three stations. The results presented in this paper refer to data during the period from 20:42 UT to 20:52 UT on September 21 (Figure 1). A ssc with a prominent negative kick (southward deflection) started simultaneously at all three stations (lower panel of Figure 1). Natural VLF emissions (chorus) were enhanced about 30 sec before the ssc onset (upper panel of Figure 1). Average emission frequency and bandwidth also increased. All of these effects of ssc-triggered VLF emissions had been discovered and interpreted in terms of a generation mechanism of chorus (Hayashi et al., 1968). In addition to the chorus emission, the VLF detector found harmonics of 60 Hz, which doubtless originated from a man-made power generation/transmission system. The main results of the observation of these harmonics of 60 Hz power are as follows:

(1) The harmonics were considerably stronger on the magnetic E-W component (the plane of the loop antenna is parallel to the local magnetic meridian plane) than on the N-S component.

(2) There is no indication that secondary VLF emissions were triggered by the harmonics.

(3) All harmonics are multiples of 180 Hz as if it were the fundamental frequency.

(4) Prior to the ssc onset, only odd harmonics of 60 Hz were detected. After onset, both odd and even harmonics occurred. Their strength changed in time concurrently with the ULF oscillations.

(5) Time changes in the harmonics were delayed in phase when compared to the time changes of the ULF oscillations. This is particularly clear in the case of the enhancement of the 3rd (180 Hz) and 6th (360 Hz) harmonics during the period 20:45:50 to 20:46:39 UT. The positive deflection in the ULF north-south component at Riverton began 15 seconds earlier, at about 20:45:35 UT. No such phase delay occurred in the case of the first enhancement of the harmonics, 20:45:00-20:45:40 UT, compared to the sharp negative deflection at



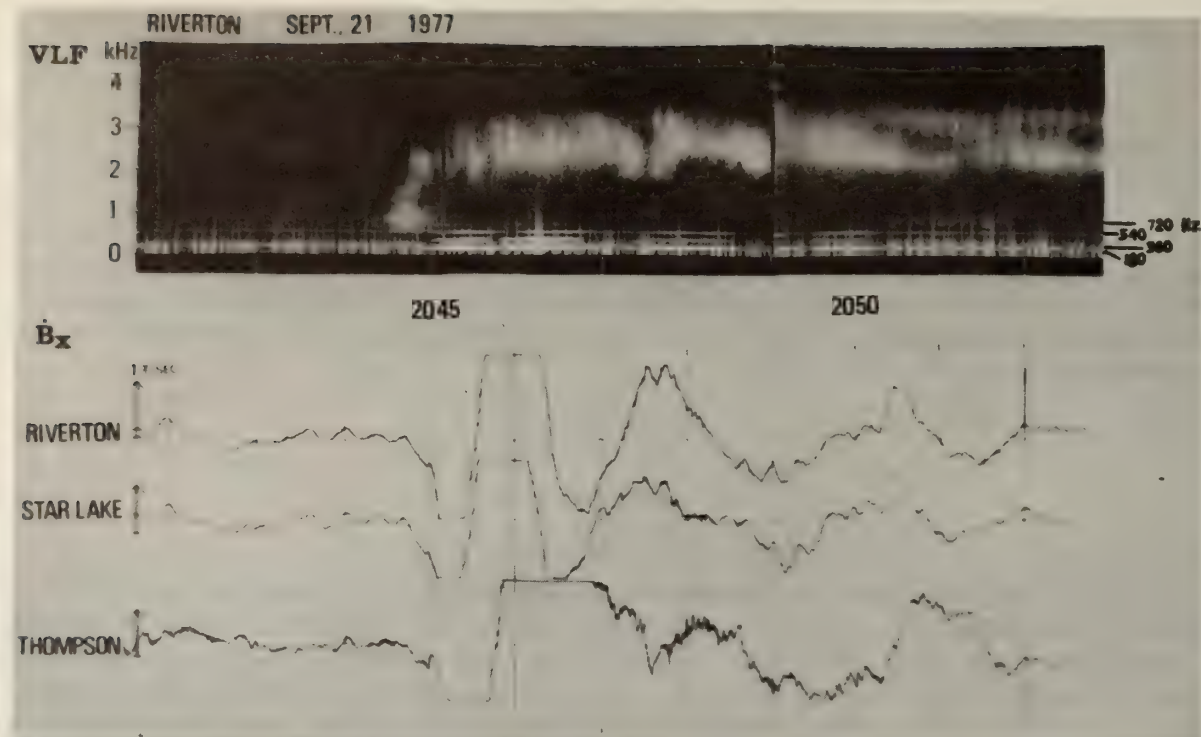


Figure 1. Natural VLF emission (chorus) and harmonics of 60 Hz man-made power activated by a magnetic disturbance (the upper panel). These were observed at Riverton, Manitoba, by a VLF receiver using a vertical loop antenna, normal to the local magnetic meridian. The ULF deflection, which began with a negative kick simultaneously (20:45 UT) at the three stations, has been identified as a ssc (the lower panel). Before the ssc onset, only odd harmonics were observed; the 3rd harmonic (180 Hz) and 9th were predominant. After the onset, both odd and even harmonics were observed. Their intensity changed concurrently with the ULF deflections, positive and negative, with little phase difference. Correlation is the best with the ULF deflections observed at Riverton.

Riverton, providing that 20:45:00 UT is assumed to be the beginning time of the negative deflection.

Results (1), (2), and (3) indicate that the harmonics came directly from a 3-phase power transmission system without following a magnetospheric path. Result (3) suggests that the power transmission line is at a sufficiently long distance from the VLF station at Riverton so that the separations between the three phase conductors of the power line are negligible. Under this condition, the local magnetic field strength at the observation site becomes proportional to the sum of the instantaneous intensities of the three currents. Most harmonics then vanish due to interference, except those that have frequencies at multiples of 180 Hz. Result (1) indicates that the power line responsible for the harmonics runs in an almost north-south direction. In fact, a major power line runs north-south about 50 km west of Riverton. However, there are several minor north-south lines nearby so that it is hard to conclude that the major power line was the responsible one. Result (2) indicates that the har-

monics did not propagate through the magnetosphere. Helliwell et al. (1975), Park (1977), and Lurette et al. (1977) discovered harmonics of 60 Hz which propagated via the magnetosphere. Those harmonics often trigger secondary emissions.

Result (4) seems to suggest that the power transmission system was affected by a solar-induced current (SIC)--a portion of telluric current shunted into a power line through neutrals of Y-connected transformers. Currents between a transformer neutral and the grounding have been actually observed. The current strength has been found to be as large as 100 amperes during a severe magnetic storm (Albertson et al., 1973). Transformer cores distort 60 Hz signals due to the nonlinearities in magnetization of ferromagnetic core materials. SICs give rise to quasi-d.c. magnetic fields in transformer cores. Without SIC, no such d.c. bias is applied to transformer cores and under such circumstances only odd harmonics are generated. Petersen (1928) has shown that not the 2nd harmonic but the 3rd is predominant. However, both odd and even harmonics are generated with a d.c. bias. Generating even harmonics with a d.c. bias is the principle of fluxgate magnetometers. The process by which harmonics are generated will be analyzed mathematically in the following section.

Concerning result (5), the delay of the enhancement of the harmonics (20:45:50-20:46:39 UT) compared to the ULF deflection at Riverton may also be attributed to the hysteresis of the transformer core, since the induced current must be reversed between the first harmonics enhancement (20:45:00-20:45:45 UT) and the one presently under consideration. The delay might also be due to a phase shift which possibly exists between telluric currents and ULF magnetic field oscillations.

### 3. ANALYSIS OF GENERATION MECHANISM OF HARMONICS

A mathematical procedure is developed in this section to determine the strength of each harmonic based on the hysteresis loop of a transformer core material. Let a closed curve in Figure 2 be an arbitrary hysteresis loop. The B-h coordinates are so chosen that the smallest and largest values of h are  $\pm h_0$  and that the origin is at the middle of the segment of the line connecting the two points on the loop that correspond to the extreme values of h. The segment represents an "average" magnetization:

$$B = \mu_r h, \quad -h_0 \leq h \leq h_0 \quad (1)$$

where  $\mu_r$  is referred to as the reversible permeability according to Bozorth (1951). Let  $B = B_1(h)$  and  $B = B_2(h)$  be the lower and upper portions of the hysteresis loop, which are connected to each other at the two points of the extreme values of h. Let us subtract the "average" magnetization from the B-h loop. The loop thus defined has the lower portion and the upper one, connected to each other at the two points on the h-axis,  $(-h_0, 0)$  and  $(h_0, 0)$ . The lower and upper magnetization curves are then expressed as follows:

$$b^{(1)}(h) = B_1(h) - \mu_r h \quad (h \text{ increasing}) \quad (2)$$

$$b^{(2)}(h) = B_2(h) - \mu_r h \quad (h \text{ decreasing}) \quad (3)$$

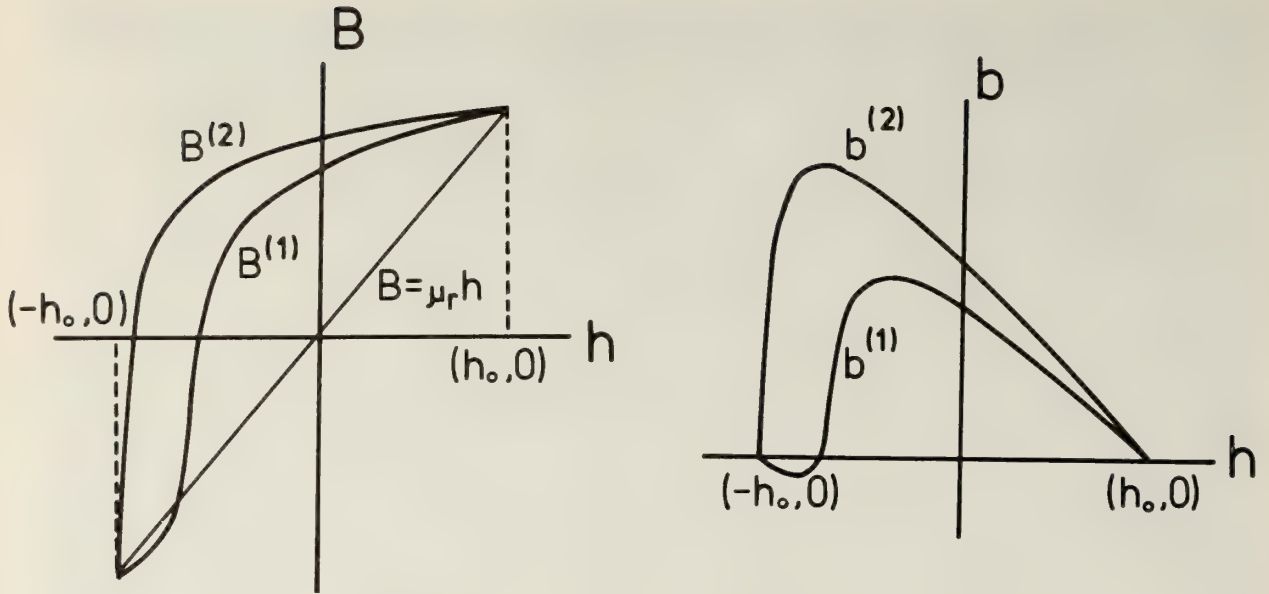


Figure 2. Reference diagrams for the mathematical analysis of harmonics.

Now, suppose that  $h$  changes with time  $t$  sinusoidally as follows,

$$h = h_0 \cos t \quad (4)$$

Then, the magnetic flux  $b$ , represented by the  $b^{(1)}$  plus  $b^{(2)}$  loops, becomes a univalent and periodic function of time and can be expressed as a Fourier series,

$$b(t) = b_0 + \sum_{n=1}^{\infty} (b_n^c \cos nt + b_n^s \sin nt) \quad (5)$$

The Fourier coefficients are determined as follows,

$$b_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} b(t) dt = \frac{1}{2\pi} \int_{-\pi}^0 b^{(1)}(t) dt + \frac{1}{2\pi} \int_0^{\pi} b^{(2)}(t) dt \quad (6)$$

and for  $n = 1, 2, \dots$

$$\left. \begin{matrix} b_n^c \\ b_n^s \end{matrix} \right\} = \frac{1}{\pi} \int_{-\pi}^{\pi} b(t) \begin{matrix} \cos \\ \sin \end{matrix} nt dt = \frac{1}{\pi} \int_{-\pi}^0 b^{(1)}(t) \begin{matrix} \cos \\ \sin \end{matrix} nt dt + \frac{1}{\pi} \int_0^{\pi} b^{(2)}(t) \begin{matrix} \cos \\ \sin \end{matrix} nt dt \quad (7)$$

Putting

$$A(t) = \frac{b^{(2)}(t) + b^{(1)}(t)}{2} \quad (8)$$

and

$$D(t) = \frac{b^{(2)}(t) - b^{(1)}(t)}{2} \quad (9)$$



we have

$$b^{(1)}(t) = A(t) - D(t) \quad (10)$$

and

$$b^{(2)}(t) = A(t) + D(t) \quad (11)$$

Then, using  $A(t)$  and  $D(t)$ , the Fourier coefficients can be expressed as follows:

$$b_0 = \frac{1}{2\pi} \left[ \int_{-\pi}^{\pi} A(t) dt + \int_0^{\pi} \{D(t) - D(t-\pi)\} dt \right] \quad (12)$$

and for  $n = 1, 2, \dots$

$$\left. \begin{matrix} b_n^c \\ b_n^s \end{matrix} \right\} = \frac{1}{\pi} \left[ \int_{-\pi}^{\pi} A(t) \begin{matrix} \cos \\ \sin \end{matrix} nt dt + \int_0^{\pi} \{D(t) - (-1)^n D(t-\pi)\} \begin{matrix} \cos \\ \sin \end{matrix} nt dt \right] \quad (13)$$

In order to perform the integrations in the two previous equations, the expressions of  $A(t)$  and  $D(t)$  have to be obtained. Regarding the lower and upper portions of the b-h hysteresis loop as a function of  $h$ ,  $A$  and  $D$  can also be considered as functions of  $h$

$$A(h) = \frac{b^{(2)}(h) + b^{(1)}(h)}{2} \quad (14)$$

$$D(h) = \frac{b^{(2)}(h) - b^{(1)}(h)}{2} \quad (15)$$

We then express  $A(h)$  and  $D(h)$  as a Fourier series, respectively, as follows:

$$\left. \begin{matrix} A(h) \\ D(h) \end{matrix} \right\} = \left\{ \begin{matrix} a_0 \\ d_0 \end{matrix} \right\} + \sum_{k=1}^{\infty} \left[ \left\{ \begin{matrix} a_k^c \\ d_k^c \end{matrix} \right\} \cos \left( \frac{\pi k}{h_0} h \right) + \left\{ \begin{matrix} a_k^s \\ d_k^s \end{matrix} \right\} \sin \left( \frac{\pi k}{h_0} h \right) \right] \quad (16)$$

Using the well-known formulae,

$$\cos(\pi k \cos t) = J_0(\pi k) + \sum_{m=1}^{\infty} (-1)^m 2J_{2m}(\pi k) \cos 2mt \quad (17)$$

and

$$\sin(\pi k \cos t) = \sum_{m=1}^{\infty} (-1)^{m-1} 2J_{2m-1}(\pi k) \cos(2m-1)t \quad (18)$$

the right-hand side of equation (16) can be rewritten as a Fourier series with  $t$  as an independent variable. In this way, we obtain the following expressions,

$$\left. \begin{matrix} A(t) \\ D(t) \end{matrix} \right\} = \left\{ \begin{matrix} A_0 \\ D_0 \end{matrix} \right\} + \sum_{\ell=1}^{\infty} \left\{ \begin{matrix} A_{\ell}^c \\ D_{\ell}^c \end{matrix} \right\} \cos \ell t \quad (19)$$

where

$$\left\{ \begin{matrix} A_0 \\ D_0 \end{matrix} \right\} = \left\{ \begin{matrix} a_0 \\ d_0 \end{matrix} \right\} + \sum_{k=1}^{\infty} \left\{ \begin{matrix} a_k^c \\ d_k^c \end{matrix} \right\} J_0(\pi k) \quad (20)$$

and for  $\ell = 2m = \text{even}$ ,

$$\left. \begin{matrix} A_{2m}^c \\ D_{2m}^c \end{matrix} \right\} = 2(-1)^m \sum_{k=1}^{\infty} \left\{ \begin{matrix} a_k^c \\ d_k^c \end{matrix} \right\} J_{2m}(\pi k) \quad (21)$$

and for  $\ell = 2m-1 = \text{odd}$ ,

$$\left. \begin{matrix} A_{2m-1}^c \\ D_{2m-1}^c \end{matrix} \right\} = 2(-1)^{m-1} \sum_{k=1}^{\infty} \left\{ \begin{matrix} a_k^s \\ d_k^s \end{matrix} \right\} J_{2m-1}(\pi k) \quad (22)$$

Calculation of the right-hand sides of equations (12) and (13) can be carried out with the help of equation (19). We finally obtain the expressions for the Fourier coefficients as follows,

$$b_0 = A_0 \quad (23)$$

$$b_n^c = A_n^c \quad (24)$$

and

$$b_n^s = \frac{2}{\pi} \{1 - (-1)^n\} D_0 + \frac{2}{\pi} \sum_{\ell=1}^{\infty} \{1 - (-1)^{n+\ell}\} \frac{n}{n^2 - \ell^2} D_{\ell}^c \quad (25)$$

On the right-hand side of equation (25), the summation with respect to  $\ell$  should be made, skipping the case where  $\ell = n$ .

The e.m.f. of the fundamental mode and of each higher harmonic is proportional to the following quantities:

$$E(1) = \sqrt{(\mu_r + b_1^c)^2 + (b_1^s)^2} \quad (\text{for the fundamental mode}) \quad (26)$$

$$E(n) = n \sqrt{(b_n^c)^2 + (b_n^s)^2} \quad (\text{for } n = 2, 3, \dots) \quad (27)$$

In the case where the B-h hysteresis loop is point-symmetry in shape, it can be shown that all even harmonics vanish. In this case of point-symmetry, we see by inspection that  $b(t)$  takes the identical form for each half-period except for sign, which alternates for every half period,

$$b(t) = -b(t+\pi) \quad (28)$$

Then the Fourier coefficients in equation (7) become

$$\left. \begin{matrix} b_n^c \\ b_n^s \end{matrix} \right\} = \frac{1}{\pi} \int_0^{\pi} b(t) \begin{cases} \cos \\ \sin \end{cases} nt \, dt \quad (29)$$

Non-existence of even harmonics can also be proved as follows. Because of the point-symmetry,

$$b^{(1)}(h) = -b^{(2)}(-h) \quad (30)$$

and consequently,

$$A(-h) = -A(h) \quad \text{and} \quad D(-h) = D(h) \quad (31)$$

Therefore,  $a_k^C$  and  $d_k^S$  in equation (16) vanish. Equations (21) and (22) then show

$$A_n^C = 0 \quad (n = \text{even}) \quad (32)$$

$$D_n^C = 0 \quad (n = \text{odd}) \quad (33)$$

We see then from equation (24) that  $b_n^C$  vanishes for  $n = \text{even}$ . From equation (25), we see that  $b_n^S$  vanishes also for  $n = \text{even}$ . In equation (25), only even terms,  $\ell = \text{even}$ , have to be taken into the summation, since all the odd terms vanish because of equation (33). Then, we see that every term in equation (25) vanishes for  $n = \text{even}$ , since  $n + \ell = \text{even}$ .

When the B-h hysteresis loop is symmetric with respect to the h-axis, it can be seen that all odd harmonics vanish, instead of the even harmonics.

The Fourier components were computed with respect to the two hysteresis loops shown in Figure 3. The hysteresis loop for the non-biased case, which is of a familiar shape, has a point of symmetry. The loop for the d.c. biased case lacks point-symmetry in general (Bozorth, 1951). With the Fourier coefficients computed, a quantity,  $E(n)$ , defined by equation (27), was calculated for  $n = 2, 3, \dots$ . The results are plotted in Figure 3. In each case, the quantities are normalized to that of the third harmonic. The hysteresis loops given in Figure 3 might not be typical of transformer core materials. However, they serve the purpose to show how different harmonics are generated dependent on the shape of a hysteresis loop. It is seen that in the case of a d.c. bias not only the odd harmonics of 60 Hz but also the even harmonics exist. If one measures the current in each individual conductor of a 3-phase power line, it should contain every harmonic of 60 Hz. This is not the case, however, if the magnetic field is detected using a VLF receiver at a far distance from the power line. Magnetic fields due to 3 phase currents vanish due to interference at most harmonics except for the frequencies at multiples of 180 MHz.

#### 4. CONCLUSIONS

The observations show that SICs seem to control harmonics of 60 Hz by giving rise to a d.c. bias to transformer core materials. Simultaneous observations should be made of geomagnetic variations, telluric currents, SICs and harmonics of 60 Hz man-made power in order to confirm the interpretation. The harmonics seem to cause considerable trouble to power systems. A general formula has been obtained to calculate the strength of each harmonic, based on the hysteresis loop of a transformer core material. It makes it possible to estimate the strength of harmonics, provided that the magnetic property of the transformer core materials is known along with the electromagnetic parameters of the transformer/power line system.



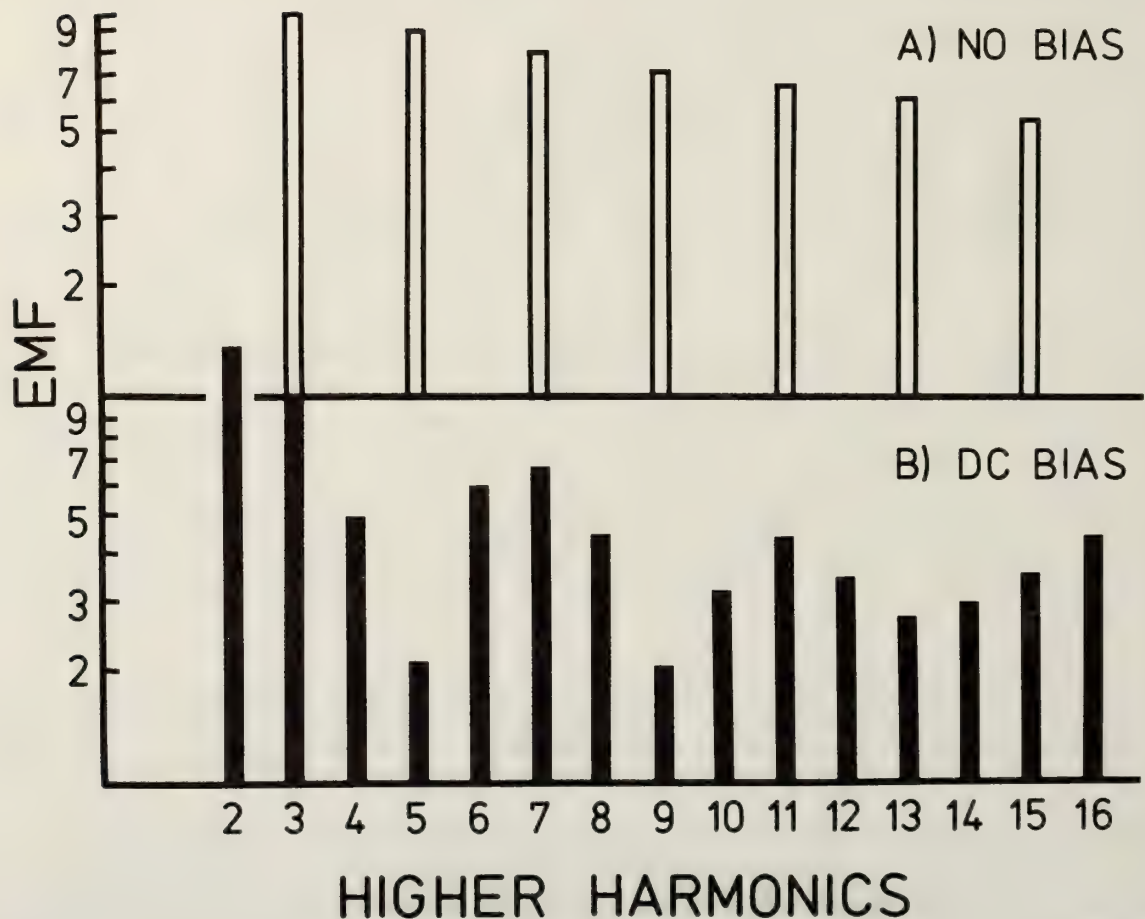
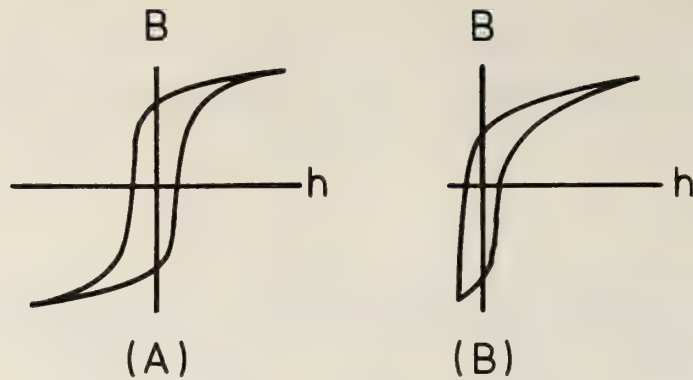


Figure 3. Magnetic flux density,  $b$ , in a ferromagnetic material excited by a sinusoidal magnetic field,  $h$ . Harmonics are generated due to the non-linearities of a hysteresis loop. On the ordinate, the amplitude of each harmonic times the frequency is plotted on a logarithmic scale. The quantity is proportional to e.m.f. The upper panel corresponds to a hysteresis loop with no d.c. bias whereas the lower panel is for a case biased by a d.c. magnetic field. In each case, e.m.f.'s are normalized to the value of the third harmonic. Only odd harmonics are generated if there is no d.c. bias applied. With a d.c. bias, however, both odd and even harmonics are excited.

## Acknowledgments

The authors sincerely thank the National Research Council of Canada, Atmospheric Environment Service Canada, Manitoba Hydro and the University of Manitoba whose assistance made the field operation possible. The authors appreciate also the assistance given by the Defence Research Establishment Pacific and the Victoria Geophysical Observatory, Earth Physics Branch of the Department of Energy, Mines and Resources. The instruments that the authors used on loan from these institutions were essential for the field work and for the analysis of the data. This research was supported by grants from the National Research Council of Canada, Arctic and Alpine Research Committee of the University of British Columbia, Ministry of Education of Japan, and the Japan Society for Promotion of Science.

## REFERENCES

- Albertson, V. D., J. M. Thorson, R. E. Clayton, and S. C. Tripathy (1973): Power apparatus and systems. IEEE Trans., PAS-92, 1031, 471-477.
- Bozorth, R. M. (1951): Ferromagnetism. New York, Van Nostrand Co.
- Hayashi, K., T. Oguti, T. Watanabe, K. Tsuruda, S. Kokubun, and R. E. Horita (1978): Nature, 275:627-629.
- Hayashi, K., S. Kokubun, and T. Oguti (1968): Rept. Ionos. Space Res. Japan, 22:149-160.
- Helliwell, R. A., J. P. Katsufakis, T. F. Bell, and R. Raghram (1975): J. Geophys. Res., 80:4249-4258.
- Lurette, J. P., C. G. Park, and R. A. Helliwell (1977): Geophys. Res. Lett., 4:275-278.
- Park, C. G. (1977): J. Geophys. Res., 82:3251-3260.
- Petersen, E. (1928): Bell Systems Tech. Journ., 7:762-796.

INDUCED ELECTRIC CURRENTS IN THE ALASKA OIL PIPELINE  
MEASURED BY GRADIENT, FLUXGATE, AND SQUID MAGNETOMETERS

Wallace H. Campbell  
U.S. Geological Survey  
P.O. Box 25046 MS 964  
Denver Federal Center  
Denver, Colorado 80225  
and  
James E. Zimmerman  
National Bureau of Standards  
Boulder, Colorado 80303

Fluxgate magnetometers in a gradient alignment and a gradient SQUID magnetometer were used to determine the current induced in the Alaska oil pipeline during a period of geomagnetic disturbance. The measurements compared favorably to each other and to the nearby current determinations using a shunt connected directly to the pipe.

## INTRODUCTION

Electric current flowing in a linear conductor produces a magnetic field, encircling the conductor, which diminishes in intensity as a function of the distance. In this paper we describe measurements of the gradient of the magnetic field perpendicular to the axis of the Alaska pipeline that were used to infer the magnitude of the electric current induced in the pipe from auroral zone, ionospheric current sources.

The 1280-km-long Alaska oil pipeline bisects the state of Alaska as a conductor that is electrically grounded throughout its length. The end-to-end resistance is less than 10 ohms. The route of the pipeline extends under the auroral zone, which has an active region maximum that occurs at latitudes usually between the center and northern terminus of the pipeline (Campbell, 1978). During active auroral periods strong electric currents, flowing in the ionospheric region of the luminous aurora near 100-km altitude, give rise to large fluctuations of the geomagnetic field at the Earth's surface beneath the aurora of hundreds and even thousands of gammas (1 gamma = 1 nanotesla). The changing magnetic fields induce large electric fields and currents in the relatively poorly conducting earth. If the Earth's surface were homogeneous in conductive properties, the pattern of this induced current at the surface would mirror that of the ionospheric



source and show a preference for the geomagnetic east-west direction. Earth crustal inhomogeneities modify this pattern of induced currents. The long, highly conducting pipe, which is electrically connected to the earth at regular intervals along the below-surface sections, has been shown (Campbell, 1978) to carry an electric current of a size just necessary to eliminate the induced electric field parallel to the pipe and Ohm's law applies:

$$I = kE_y \quad , \quad (1)$$

where  $I$  is the current (amperes) in the pipe,  $E_y$  (volts/meter) is the electric field along the pipeline, and  $k$  (meters/ohm) is the conductivity of the pipe. Appendix B of an earlier paper by Campbell (1978) gives the justification of this equation.

The horizontal electric field,  $E_y$ , is dependent upon the amplitude of the perpendicular horizontal, geomagnetic field variation  $B_x$  (gammas), the period of this inducing variation  $T$  (minutes), and the fixed conductivity structure of the immediate earth subsurface, as first noted by Cagniard (1953). To a first approximation we may write,

$$I = k'B_x T^{-0.5} \quad , \quad (2)$$

with  $k'$  a constant which includes the effective earth conductivity and therefore varies with differing measurement-site geology.

The magnetic field encircling a linear electric current of the type induced along the Alaska pipeline is

$$B_z = 10^9 \mu_0 I / 2\pi R = (200 \text{ ohm s/m}) I / R \quad , \quad (3a)$$

or

$$B_z = 200 I / R \quad (3b)$$

where  $R$  is the distance from the center of the pipe, and  $\mu_0 = 4\pi \times 10^{-7}$  ohm s/m is the permeability of free space. With  $I$  in amperes and  $R$  in meters,  $B_z$  will be given in gammas. Figure 1 illustrates the decrease of field with distance for a 1 ampere current. The derivative of Equation 3, rewritten to obtain the current, is

$$I = - 5 \times 10^{-3} R^2 dB_z / dR \quad , \quad (4)$$

where  $dR$  (meters) is the increment in distance directed radially from the pipeline and  $dB_z$  (gammas) is the corresponding increment in magnetic field in a direction that would encircle the pipeline axis. For our measurements  $dR$  was in the horizontal plane containing the pipeline axis and  $dB_z$  was the corresponding change in vertical component of the magnetic field. Because only the time varying part of the inducing geomagnetic field gives rise to the current fluctuations, only the fluctuations of field gradient,  $dB_z/dR$ , are measured; the local, constant gradients of field, due to the magnetization of the iron in the pipe and of the local geological materials, are considered to be only the constant baseline level from which such fluctuations are determined. It was assumed that the field of the induced magnetization of the pipe would be much smaller than the field of

the induced current in the pipeline.

Although the induced currents may become a problem for the Alaska pipeline metering systems that are unprotected for the large surges (over 1000 amperes) that can occur during the great geomagnetic storms, engineering solutions to such problems can be readily determined. The principal concern presently is the contribution to corrosion that can occur at the small holes in the pipe coating for underground sections when that current flows between the pipe and ground. Presently, exploratory studies are underway to determine the favored regions for such current flow.

In this paper we describe the application of the gradient method for surveying the current in the Alaska pipeline. Two systems were tested at the Fox Junction site ( $64.95^{\circ}$  N.,  $152.3^{\circ}$  E.), about 10 km NNE of Fairbanks, Alaska, for a brief period in August, 1978. In one method the Z (vertical) component of magnetic field was measured with two fluxgate magnetometers at two fixed distances in a horizontal line perpendicular to the pipeline axis. In the other method a specially designed gradient SQUID (acronym for Superconducting Quantum Interference Device) magnetometer for the Z component of field was used at a distance from the pipeline corresponding to that of the close fluxgate magnetometer. The induced current determinations were compared to each other and to earth electric field measurements available from a nearby observatory. Finally, the currents determined by the magnetometers were compared to nearby values obtained by the pipeline operations personnel using a current shunt connected directly to the pipe.

## OBSERVATIONS

### A. Fluxgate Operation

The two fluxgate magnetometers (A and B) were located in a line perpendicular to the pipeline axis at 10 and 50 meters from the pipe center at a location where the pipeline was straight for a distance that was quite long with respect to 50 m. Figure 1 illustrates the positioning of the magnetometers relative to the induced-current field given in Equation 3. For this fixed 40-m separation ( $\Delta R$ ) Equation 4 becomes

$$I = 6.25 \times 10^{-2} \Delta B_z, \quad (5)$$

where  $I$  is the pipeline current (amperes) and  $\Delta B_z$  (gammas) is the difference in amplitude of vertical field for magnetic fluctuations measured by the two magnetometers. Forming  $\Delta B_z$  in this fashion eliminates the contribution to  $B_z$  from the natural ionospheric source field whose scale size is such that the contribution to field variations would be the same at the A and B magnetometer locations. Figure 2 illustrates the field observed by the two fluxgate magnetometers. Note that the field from the induced pipeline current dominates the field appearance at the near (10-m) location whereas the distant (50-m) location record shows the admixture of ionospheric source field and pipeline current field.

Induced current variations due to local geologic anomalies also contribute to the Z component field measurements. It was assumed that the



Figure 1.--Nearby field  $B_z$  (gammas) from 1 am-pere pipeline current at distance  $R$  (meters) from the pipe center. The location of the magnetometers is indicated along with the equivalent gradient location for the fluxgates. 1 gamma = 1 nanotesla.

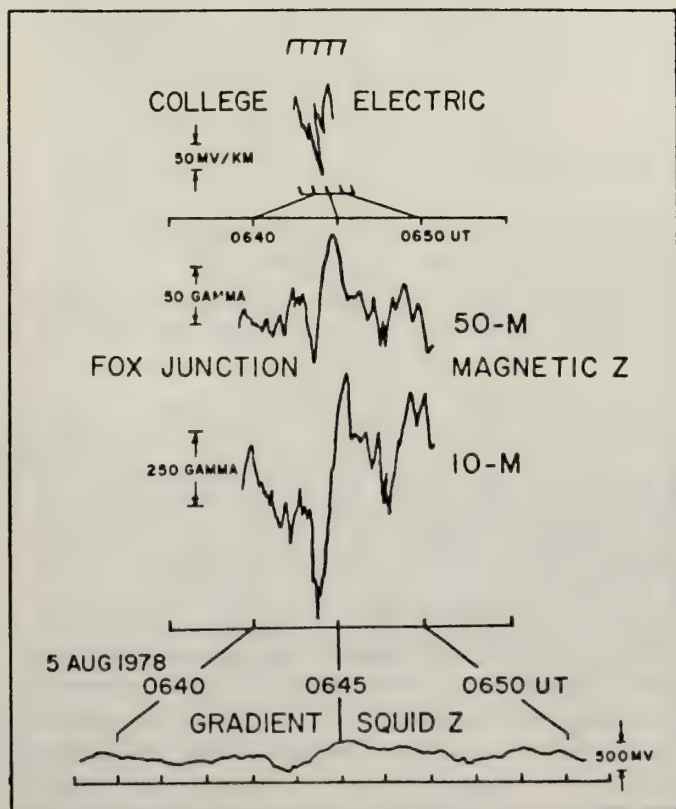
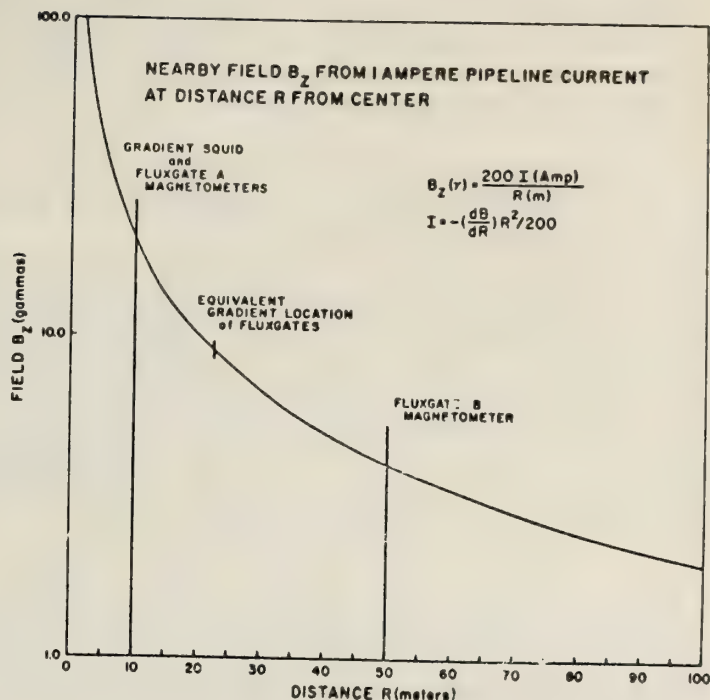


Figure 2.--Illustration of a field fluctuation recorded simultaneously by the College electric potential probes (top), Fox Junction fluxgate magnetometers at 50 m and 10 m locations (center) and the Fox Junction gradient SQUID magnetometer at the 10 m location (bottom). The sample records are shown for the period 0640 to 0650 UT on 5 August 1978.



conductivity and scale of such crustal conductivity anomalies compared to the high conductivity and small pipeline diameter was such that the Z component contribution was essentially the same for the two magnetometers separated by 40 meters and this contribution was removed by the subtraction process.

The pipeline axis at Fox Junction was about 2.8 meters above the ground so a small trigonometric correction was made to adjust for the ground level plane of the magnetometers.

A small error was introduced by the method of scaling which depended upon a determination of the difference between the maximum and minimum levels of field pulsation (e.g., that oscillation near 0645 UT in Figure 2). The pipeline current results from an interaction due to variations in the horizontal geomagnetic field component. The two-magnetometer measurement subtraction would remove the common geomagnetic source Z component of field if a running difference were being recorded. This vertical component may or may not be in phase with the horizontal variation depending upon the ionospheric current source configuration and location with respect to the earth surface observer. In scaling the two records, pulsations extremes were matched which may occur at slightly different times depending on the relative mixture of the two current field sources and the time accuracy in the scaling process. This shortcoming can account for some of the data variability seen in subsequent plots. Future observations will involve the direct recording of the A and B magnetometer difference after adjustment to a zero level during times of quiet geomagnetic field.

Continuous, natural electric earth potential in a north-south direction,  $E_{CO}$  ( $\mu$  V/m), and geomagnetic northward field fluctuation,  $B_x$  (gammas), measurements were available from College ( $64.87^\circ$  N.,  $212.2^\circ$  E.) about 11 km SW of the Fox Junction observation site. Our determination of the Fox Junction current,  $I_{FJ}$  was related to  $E_{CO}$  by the linear regression

$$I_{FJ} = 0.405 E_{CO} \quad , \quad (6)$$

for which the regression coefficient was 0.94 in accordance with the expectation of Equation 1. Then, using the estimate of current corresponding to  $E_{CO}$  obtained from this relationship and the simultaneous values of  $B_x$  from the College observatory for scaled variations of period T (minutes), we obtained

$$I_{FJ} = 0.65 B_x T^{-0.5} \quad (7)$$

as a relationship corresponding to Equation 2 for obtaining an estimate of the pipeline-induced electric current near Fairbanks from the observatory geomagnetic field levels. Figure 3 illustrates this formulation for a range of typical values of  $B_x$  and T.

In a very quiet geomagnetic period at a region of the pipeline route near Abercombie Creek, Valdez ( $61.10^\circ$  N.,  $213.8^\circ$  E.), where the pipeline was buried with its axis about 3.3 m below the surface, we made a brief survey of the Z component of field along a line perpendicular to the pipeline axis. Figure 4 illustrates the observed field levels from which we verified the horizontal location of the pipeline axis with respect to the

Figure 3.--Pipeline current near Fairbanks estimated from the local north-south magnetic field fluctuations  $B_x$  (gammas) for various typical periods  $T$  (minutes) as given in Equation 7 of text.

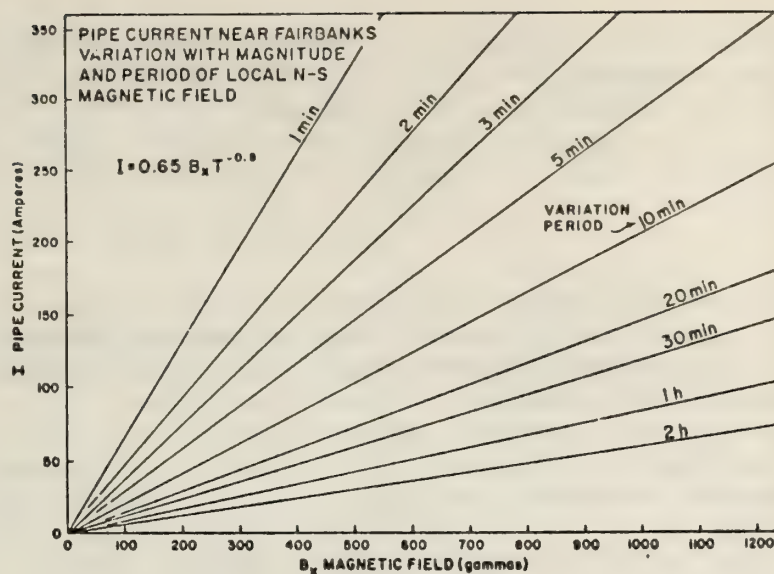
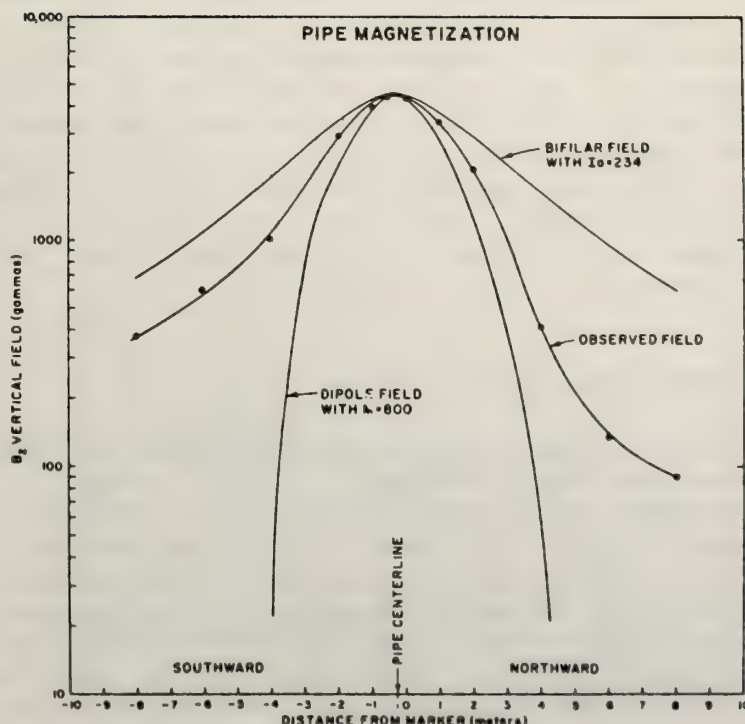


Figure 4.--Observed variation of vertical magnetic field,  $B_z$ , (scale to left) with horizontal marker-distance from the magnetized pipe buried 3.3 m below the surface; field of a dipole with moment  $M = 800$  ampere-meters<sup>2</sup> (scale to left) centered at the pipeline axis; and field of bifilar system with conductor separation,  $a$ , small with respect to buried depth and the value of  $Ia = 234$ , fixed to fit the observed field above pipe centerline.





construction contractor's markers at ground level and determined the pipe magnetization. The centerline was identified at a location only about 0.3 m south of the construction markers.

The vertical component of field  $B_z$  (gamma) due to a vertical magnetic dipole of moment  $M$  (amperes . meter<sup>2</sup>) is given by the relationship

$$B_z = 10^2 M (3 \cos^2 \theta - 1) R^{-3} \quad , \quad (8)$$

where  $R$  is the distance from the dipole center to the point of observation in meters, and  $\theta$  is the angle between a position directly above the dipole and the point of observation. At large distances from the pipe,  $3 \cos^2 \theta$  becomes zero and the field varies as  $R^{-3}$ . Near to the pipeline axis the measurements behave as a dipole field with a moment of approximately  $800 \text{ Am}^2$  as indicated in Figure 4. The fact that the observed magnetization field drops off less rapidly than that of the dipole is, in part, due to the linear extent of the pipeline. Just as a magnetic dipole may be represented by a circular loop of current, an infinite line of dipoles may be represented by a bifilar current. The field,  $B_z$  (gammas), of a bifilar current,  $I$  (amperes), system with parallel elements separated by a distance,  $a$  (meters), is given by

$$B_z = 200 I a \left( \frac{x/a + 1/2}{a_2^2} - \frac{x/a - 1/2}{a_1^2} \right) \quad , \quad (9)$$

where  $x$  is the horizontal distance from the axis to  $P$ , the point of observation;  $a_1$ , and  $a_2$  are the distances from the two bifilar conductors to  $P$ . Assuming that  $a$  is small with respect to the distance to  $P$ ,  $a_1 = a_2 = R$  so that the field changes as  $R^{-2}$ . A bifilar representation of the pipe magnetization field for  $I a = 234$  is also shown in Figure 4. The observed fields seem to indicate that the observed pipe magnetization field lies between these two models and superposed on a local spatial gradient that decreases from south to north. The field gradients due to the local geologic features and the permanent magnetization of the pipe do not vary with time and therefore do not contribute to our determinations of the field gradient from the induced pipeline current variations.

## B. SQUID Operation

The electronics, the cryostat, and the SQUID used for the present measurements were similar, in all respects but one, to the SQUID magnetometer system described by us several years ago (Zimmerman and Campbell, 1975). The only difference was that the "fractional-turn" or multihole SQUID used for the pipeline measurements was built and mounted to respond to the horizontal gradient of the vertical field, rather than to the vertical field itself (see Figure 5). With a multihole SQUID, this is accomplished simply by connecting half the holes to give positive output for an up field and the other half to give positive output for a down field. Thus, if the SQUID is mounted with the holes vertical, it responds to the change of vertical field across a horizontal diameter, a term in the field gradient of the form  $\partial B_z / \partial x$ , where  $B_z$  is the vertical field component, and



Figure 5.--Fractional-turn ("multi-hole") gradient SQUID, oriented to measure  $\partial B / \partial x$ , mounted in the  $z$  bottom of a liquid-helium cryostat. Support structure, r.f.-bias coil, coaxial line, and room-temperature electronics (not shown) is similar to that given in Zimmerman and Campbell (1975). Above left is a symbolic sketch of an elementary 2-hole SQUID connected as a gradiometer and as a magnetometer. Meaning of the symbols is given above right.

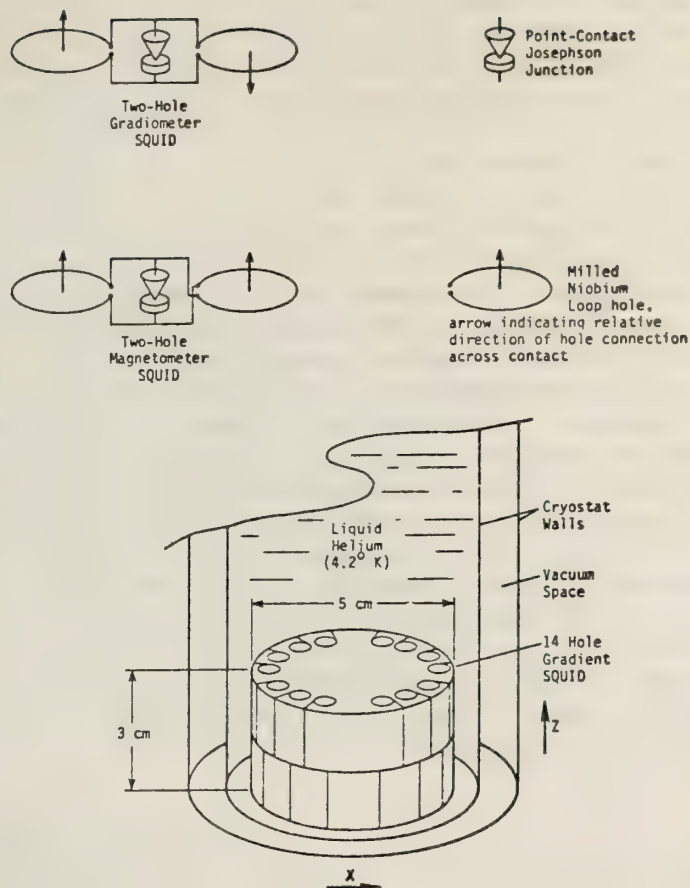
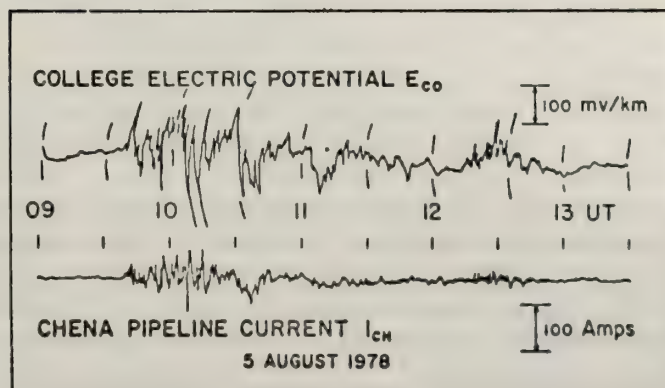


Figure 6.--Comparison of College electric potential,  $E_{CO}$  (mv/km), top trace, and Chena pipeline current  $I_{CH}$  (amperes), bottom trace, for representative interval between 0900 and 1330 UT on 4 August 1978.



the radial horizontal distance from the center of the pipe.

Any asymmetries in a gradient sensor will result in a response to field as well as to gradient. The SQUID gradiometer was trimmed by means of superconducting tabs placed empirically to correct the asymmetries, that is, to eliminate the response to uniform field. This was accomplished to an acceptable degree for the horizontal field components. However, in the time available, the response to vertical field was not reduced as much as it should have been. As a result, with the gradiometer 10 m from the pipe, its response was an algebraic sum of two terms, one proportional to the field  $B_z$  and one proportional to the gradient  $\partial B_z / \partial R$ . In order to derive the gradient term alone, we subtracted a correction derived from the field  $B_z$  as measured by fluxgate magnetometer A (see Figure 1), which was also 10 meters from the center of the pipe.

Because of the residual asymmetry of the SQUID gradiometer, and the method of correcting for it, the gradiometer measurements were, therefore, not completely independent of the fluxgate magnetometer measurements. Nevertheless, the gradient term in the gradiometer response was independent of the fluxgate magnetometers, so any gross inconsistency between the two measurement methods should still be apparent in comparing the corrected results. An example of the gradiometer response is shown at the bottom of Figure 2.

The SQUID was calibrated by recording its response to a small bar magnet of known moment at several distances  $x_1, x_2 \dots$  of the order of 1 m. The response may be written

$$r = A B_z + B \partial B_z / \partial x \quad (10)$$

where  $\partial B_z / \partial y$  for a vertically oriented magnetic dipole, along a line perpendicular to the dipole, is

$$\partial B_z / \partial x = - 3 B_z / x \quad (11)$$

Thus, placing the bar magnet at  $y_1 = 1$  m and  $y_2 = -1$  m, respectively, from the center of the SQUID, gave two responses  $r_1$ , and  $r_2$ , from which the calibration constants A and B were determined. The results were  $A^{-1} = 0.71$  gamma/mV and  $B^{-1} = 0.165$  gamma/vM.m. With the gradiometer 10 m from the pipe, the two terms in the response were the same order of magnitude.

The use of a small bar magnet at measured distances from the gradiometer proved to be a very useful indicator of the asymmetries referred to above, and could probably be used to balance the asymmetries to one part in  $10^3$  or better, which would have been good enough for the pipeline measurements at 10 m distance. The gradiometer imbalance actually was measured by the much more sensitive and time-consuming method of rotating the instrument in the Earth's field. The relative merits of the two methods are that the former can be done relatively quickly in the laboratory, while the latter requires a location free of artificial magnetic anomalies (i.e., away from the laboratory), and a moderately sophisticated mechanism for rotating the cryostat. On the other hand, the Earth's field in many locations is highly uniform and enables one to achieve an imbalance of the order of one part in  $10^6$ . ("Imbalance" here is defined as the response to uniform field compared to what it would be if both halves of the gradiometer were connected to measure sums rather than differences of fields.)

The instrument noise of the gradiometer was less than  $10^{-3}$  gamma/m for an output-averaging time of one second. A current of one ampere at a distance of 10 meters gives a gradient (Equation 4) of 2 gamma/m, far greater than the instrument noise level.

### C. Comparison of Measurements

A comparison of the currents determined by the fluxgate and cryogenic SQUID gradient methods showed

$$I(\text{fluxgate}) = 0.993 I(\text{SQUID}) \quad (12)$$

with a linear regression coefficient of 0.99, which indicates clearly that the two methods give comparable values for the induced pipeline currents.

Subsequent to the observations it was discovered that the Alyeska Pipeline Service Company had operated a current measuring shunt ( $5 \mu$  ohms) system across a 198.1 m (650 ft) spacing of the pipeline at their Chena Test site about 9.6 km SE of our measurement location. A comparison of these Chena current values,  $I_{CH}$  (amperes), and the simultaneous College electric potentials,  $E_{CO}$ , is illustrated in Figure 6. Assuming Equation 6 to obtain the equivalent Fox Junction pipeline current, a comparison of the values obtained from the two sites is given in Figure 7 from which we obtain

$$I_{FJ} = 1.05 I_{CH} \quad (13)$$

with a linear regression coefficient of 0.96.

### CONCLUDING REMARKS

The field gradient method for observing the electric currents in the Alaska oil pipeline provided consistent values for both the fluxgate and SQUID methods of observation. These currents were linearly related to the regularly measured electric and magnetic field changes at the College observatories. The determinations of pipeline current were consistent with values obtained by a direct connection, current shunt technique at a pipeline site about 9.6 km away. The gradient method of observing has the distinct advantage of portability and buried-pipe measurement capability. Field gradients due to the pipe magnetization, geological features, or ionospheric source currents do not seem to contribute a measurable error to such pipe current determinations.

The SQUID gradiometer is inherently sensitive enough to detect very small currents in a linear conductor at 10 meters, or conversely, to detect currents of one ampere or more at relatively great distances. Our experience, and conversations with SQUID gradiometer manufacturers, indicates it is fairly straightforward to achieve an imbalance less than one part in  $10^4$ , and with extreme care, one part in  $10^6$  or better.



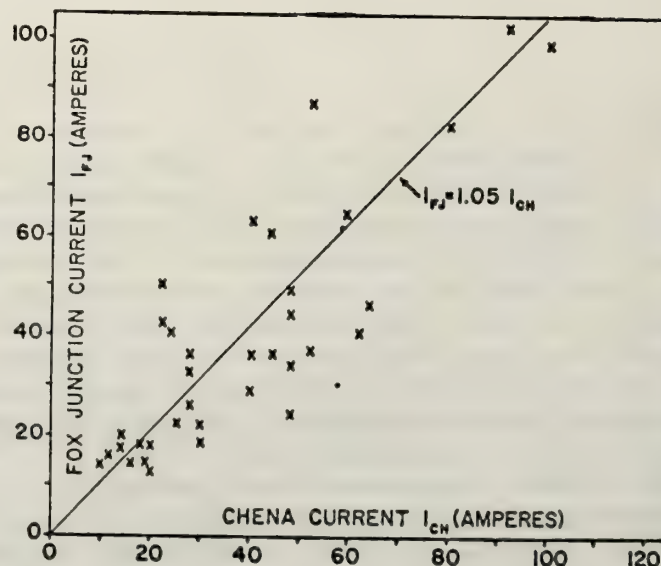
# ACKNOWLEDGMENTS

The cooperation of the Alyeska Pipeline Service Company was obtained through the help of Peter Byfield who supplied the Chena current measurements. Richard R. Heacock, at the Geophysical Institute, University of Alaska, graciously supplied copies of his earth potential records concurrent with our observations. John B. Townshend and John E. Papp of the U.S. Geological Survey College Magnetic Observatory gave us the copies of the College magnetic field records used in this study. Eugene M. Wescott of the Geophysical Institute, University of Alaska, assisted in finding a suitable observation site near Fairbanks. The authors especially thank Richard Kuberry of the U.S. Geological Survey for his help in calibrating the magnetometers.

# REFERENCES

- Cagniard, L., (1953): Basic theory of magneto-telluric method of geophysical prospecting. Geophysics, 18, 605-635.
- Campbell, W. H., (1978): Induction of auroral zone electric currents within the Alaska pipeline. Pure and Applied Geophysics, 116, 1143-1173.
- Zimmerman, J. E., and W. H. Campbell, (1975): Test of cryogenic SQUID for geomagnetic field measurements. Geophysics, 40, 269-284.

Figure 7.--Simultaneous values of Fox Junction current  $I_{FJ}$  (amperes) and Chena current,  $I_{CH}$  (amperes) for a sample period of activity from 1100 UT of 3 August to 1400 UT of 5 August 1978.



# VERY LARGE GEOMAGNETIC DISTURBANCES DURING SUNSPOT CYCLE 21 - A PREDICTION

H. H. Sargent III  
Space Environment Services Center  
325 South Broadway  
Boulder, Colo. 80303

Evidence is presented which suggests that very large geomagnetic disturbances (350 gammas or greater at an invariant magnetic latitude of 50 degrees) occur once or twice per sunspot cycle, on the average. There is also some tendency for these disturbances to group in large odd-numbered sunspot cycles similar to the current cycle, Cycle 21. No such disturbance was noted during the past cycle although a series of major solar flares was observed in August 1972. At least one very large geomagnetic disturbance is expected during the current cycle; a prediction with perhaps serious consequences for electric power companies.

## 1. INTRODUCTION

The Space Environment Services Center (commonly referred to as the Solar Flare Forecast Center) was established in 1965, and has served since then as a collection point for real-time solar data. In August 1972, a series of major solar flares occurred which impressed solar forecasters as the most energetic events seen in the short history of the Forecast Center. The Public Information Office of NOAA's Environmental Research Laboratories went into action; and soon, newspapers around the World informed readers that one of the most energetic series of solar flares ever observed was in progress. Great solar flares have occurred in the past, but none have had such immediate and widespread attention from the public. After reading the collection of news clippings reproduced by McKinnon (1972), one is left with the impression that the August 1972 series of solar events was one of the most spectacular of this century.

By early 1969, a number of electric power operating companies and organizations in the United States and Canada had installed recording dc ammeters on selected power transformers to measure quasi-dc currents caused by geomagnetic storms. This effort was part of the Solar Magnetic Disturbances Research Project sponsored by the Edison Electric Institute. A total of 64 recording locations are listed by Albertson, Thorson and Miske (1974b). While most of these recorders had been deactivated by August 1972, the EEI project resulted in the best documentation ever achieved of electric power system disturbances over a large geographical area. This was due not only to instrumentation; but also,

to increased awareness of these problems by power industry personnel. In recounting the effects of the August 1972 flare-induced geomagnetic storms, Albertson and Thorson (1974a) reported numerous electric power system disturbances, but no widespread outages or major system stability problems.

The concern addressed in this paper is that interests acquainted with both the news accounts of the August 1972 solar events and results such as Albertson and Thorson's would conclude that the electric power distribution systems of the United States and Canada had survived the worst geomagnetic storms of this century. It can easily be shown that this is not the case.

## 2. DISCUSSION

Figure 1 shows that the distribution of moderate and strong electric power system disturbances on August 4, 1972 was generally limited to the area defined as the southward extent of the aurora, given that  $K_p$  briefly reached a value of 9. The curves in Figure 1, which show the subpoints of the southern edges of the aurora for different values of  $K_p$ , are due to Sprague (1968). In general, the auroral oval expands equatorward as geomagnetic activity increases. The close relation between temporal variations of the size of the auroral oval and the magnitude of geomagnetic activity has recently been confirmed by Sheely (1978) using DMSP satellite observations of the aurora. In fact, the southernmost ground-based observation of aurora in August 1972, as

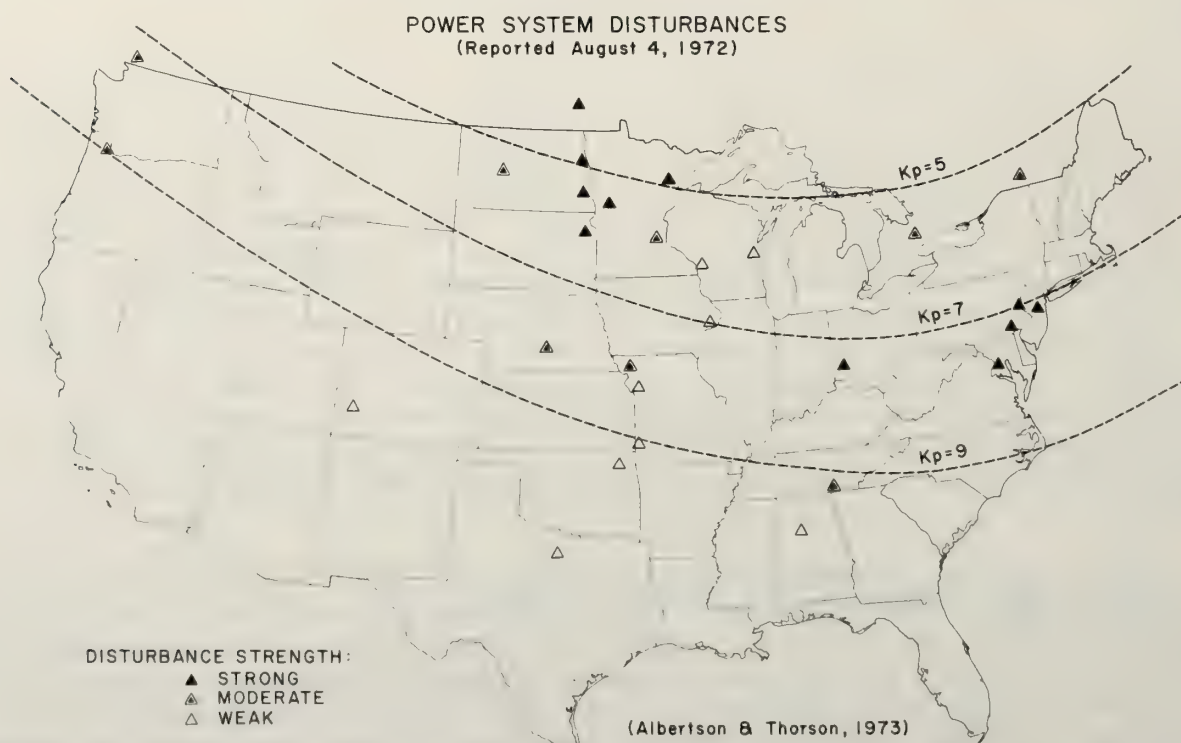


Figure 1. Map of the United States showing the location of power system disturbances on August 4, 1972. Dashed lines indicate the average southward limits of the aurora for three values of  $K_p$ , the planetary magnetic index.



reported in Sky and Telescope magazine, was in the state of Kentucky (close to the  $K_p=9$  line in Figure 1). If the southernmost extent of the aurora defines the area where serious disturbances to electric power systems may occur, consider the work of Chapman (1957). He noted aurora observations in Florida, Cuba, Hawaii and several other decidedly tropical locations and even commented on the sighting of aurora in Singapore (1 degree north of the Equator) during the great geomagnetic storm of September 25, 1909!

In order to examine the rate of occurrence of very large geomagnetic disturbances, the aa-index of Mayaud (1973) was sorted and the results plotted in Figure 2. The aa-index is published in half-day values since 1868. Since an aa-index average value of 50 gammas or more (at an invariant magnetic latitude of 50 degrees) over a 24-hour period conforms, more or less, to the Space Environment Services Center's working definition of a major geomagnetic storm, half-day values of 100 gammas or more were chosen to indicate those periods when major geomagnetic storm conditions certainly prevailed. Such half-day periods are referred to here as "high-category major geomagnetic storm" periods. The open bar graph in Figure 2 shows the total number of such periods for each year since 1878. The shaded bar graph shows the total number of separate storms during these same years (storms are often comprised of several consecutive high-category major geomagnetic storm periods). The connected points represent the annual mean sunspot number for each year since 1878.

In the course of this work, the term "super storm" was coined and applied to those half-day aa-index periods when the index reached or exceeded 350 gammas. These super storms are indicated in Figure 2 by heavy vertical lines.

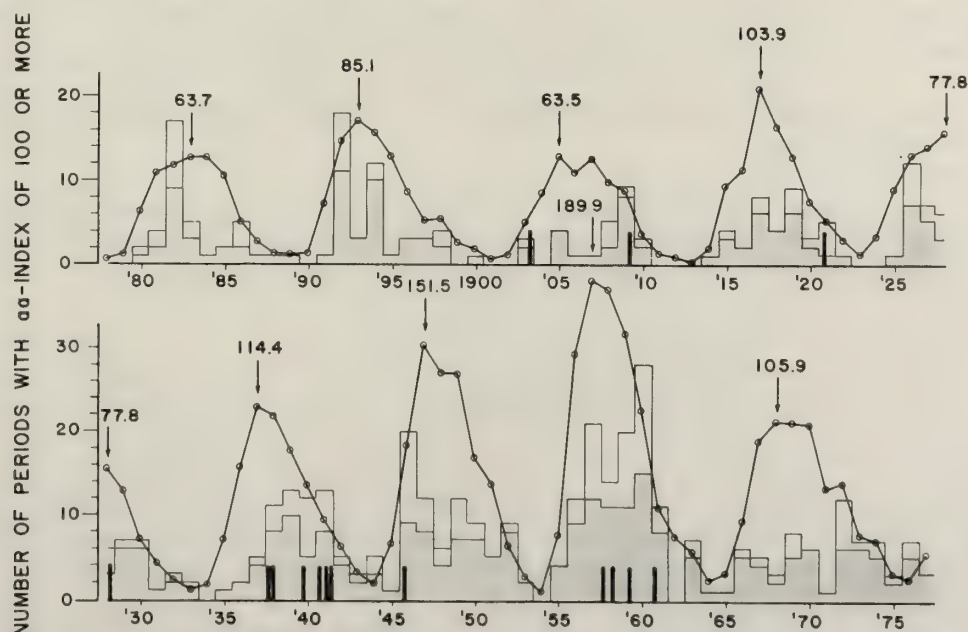


Figure 2. Bar chart showing the number of half-day periods each year when the aa-index equalled or exceeded 100 gammas. Shaded levels indicate the number of separate high-category major geomagnetic storms in a given year (some storms involve several consecutive half-day periods). Heavy vertical lines indicate super storms (where the half-day value equalled or exceeded 350 gammas). Annual mean sunspot numbers are also shown with maximum values specified for each cycle.

These are seen to have occurred some 17 times during the past nine sunspot cycles. There is reasonable agreement between this rate of occurrence and the estimated frequency of aurora observations, at low latitudes, published by Vestine (1944). In addition, most of the low latitude aurora observations discussed by Chapman are also identified with super storms. On average, one or two of these super storms can be expected during each sunspot cycle; however, there does appear to be a tendency for these storms to group in the larger odd-numbered sunspot cycles. Of special note is the September 25, 1909 super storm during which the half-day aa-index reached a value of 546 gammas! By comparison, the reading on August 4, 1972 was 220 gammas. Siscoe (1976) has shown that in one hundred sunspot cycles, one value equal to, or in excess of, 679 gammas may be expected!

Even if no super storms are observed during the current sunspot cycle, there is cause for concern. The vulnerability of electric power distribution systems to solar flare-induced geomagnetic disturbances has probably increased substantially since August 1972 with the proliferation of interconnecting ties and the reduction of spinning reserve. Albertson, Thorson, Clayton and Tripathy (1973) asserted that the trend toward longer lines and interconnections over large geographical areas would lead to further problems.

### 3. CONCLUSIONS

3.1 The events of August 1972 were not particularly noteworthy in comparison with geomagnetic disturbances noted in previous sunspot cycles.

3.2 One or more super storms (350 gammas or more) can be expected during the next four or five years.

3.3 Even if the geomagnetic storms of Cycle 21 are no stronger than those observed in August 1972, electric power disruptions are a possibility due to the increased vulnerability of electric power distribution systems.

### REFERENCES

- Albertson, V. D., Thorson, J. M. Jr., Clayton, R. E., and Tripathy, S. C. (1973): Solar-induced currents in power systems: causes and effects. IEEE Transactions on Power Apparatus and Systems, PAS-92:471.
- Albertson, V. D., and Thorson, J. M. Jr. (1974a): Power system disturbances during a K-8 geomagnetic storm: August 4, 1972. IEEE Transactions on Power Apparatus and Systems, PAS-93:1025.
- Albertson, V. D., Thorson, J. M. Jr., and Miske, S. A. Jr. (1974b): The effects of geomagnetic storms on electrical power systems. IEEE Transactions on Power Apparatus and Systems, PAS-93:1030.

- Chapman, S. (1957): The aurora in middle and low latitudes. Nature, 197:7.
- Mayaud, P. N. (1973): A hundred year series of geomagnetic data 1868 - 1967, IAGA Bulletin No. 33, Meppel, Holland.
- McKinnon, J. A. (1972): August 1972 solar activity and related geophysical effects, NOAA Tech. Memo. ERL SEL-22, Space Environment Laboratory, Boulder, Colo. 80303.
- Sheeley, N. R. Jr. (1978): The equatorward extent of auroral activity during 1973 - 1974. Solar Physics, 58:405.
- Siscoe, G. L. (1976): On the statistics of the largest geomagnetic storms per solar cycle. Journal of Geophysical Research, 81:4782.
- Sprague, G. C. (1968): Visual studies of the aurora. Sky and Telescope, 35:346.
- Vestine, E. H. (1944): The geographic incidence of aurora and magnetic disturbance, northern hemisphere. Terrestrial Magnetism and Atmospheric Electricity, 49:77.



# POTENTIAL PREDICTION NEEDS IN SUPPORT OF ENERGY SYSTEMS

Richard L. Blake

MS 436, University of California, Los Alamos Scientific Laboratory  
Los Alamos, New Mexico 87545

Utilization of existing and potential new energy systems involves interactions between these systems and the natural environment. Basic research is needed to define the interaction processes of significance. Thereafter long range solar-terrestrial predictions will be important to national policy on proper control of energy systems.

## 1. INTRODUCTION

Energy supply and distribution systems for the United States are the responsibility of the Department of Energy (DOE). In recent years it has become apparent that the most expeditious and economic means of producing the distributing energy are often at odds with the responsibility of all citizens and government to protect our natural environment. Part of the solution to this conflict lies in the improvement of our understanding of the global environment so that consequences of proposed national policies can be anticipated. Solar-terrestrial processes are part of the natural environment and these processes can indeed influence energy systems.

How will energy production, transmission and control be influenced by as well as influence the natural environment over both short and long periods? This question, legally imposed by Congress, must be answered by government agencies and organizations involved in energy systems. The response, called an environmental impact statement, is normally useless and generally ignored. We are far too ignorant of the global consequences of most energy technologies. What we do know is that "environment" certainly includes more than the immediate surroundings of an energy installation, in spite of the practice of confining impact statements to this local concept. In this report are presented a few selected examples to show that energy systems are tied to global solar-terrestrial processes. These examples have all been discussed elsewhere [see especially Lanzerotti (1978)]. They are used to emphasize the need for long-term basic research on the solar-terrestrial system.

## 2. BASIS OF PREDICTION NEEDS

Existing or potential energy systems must be assessed in regard to the availability of resources (fossil fuels, solar radiation, geothermal sources, etc.), the technology to utilize these resources and transport the energy to where it is needed, and effects on the environment. The resources are time and geography dependent while resources, technology, and environmental impact are all interdependent. Assessment of their interdependence requires geophysical, atmospheric, and solar-terrestrial data inputs.

In recent years we have learned that important variations in Earth's biosphere do arise from interactions of the near surface components with a far more extended system going out to the sun, and also from man-produced factors. Not only does solar radiation drive the circulation of the lower atmosphere with its weather patterns, but the short wavelength (ultra-violet) portion and particles are largely responsible for conditions in the high atmosphere well above the weather-dominated layer. These higher portions contain trace constituents whose importance far exceeds their fractional concentration, a prime example being ozone. The concentrations and changes in these trace constituents caused by fluctuations in solar radiation must be assessed as a baseline to compare with changes caused by energy-related technologies and other man-made injections. For example, a few years ago it was not possible to draw firm conclusions on the effects of SST exhausts on the ozone layer because there was inadequate knowledge of the stratosphere. Similarly, one cannot accurately forecast future energy needs for a model world because there is inadequate knowledge of the many factors (solar flux, volcanic dust, transportation and industrial emissions, etc.) that control the climate, or the climate feedback that must influence decisions on energy production.

Energy technologies have need of long-range predictions and understanding of solar-terrestrial processes rather than short-range predictions. The reason is that neither decisions nor technology can be changed on a national scale in days or weeks. But over months or years there is a great deal that can be done. The best example is the allocation and distribution of energy resources in response to regional changes in weather patterns or climate. If the sun has a predictable influence on climate it is a small effect that is ignorable regarding the total supply of energy, but it could have a large effect on regional demand for energy. And if the solar activity cycle has an amplitude modulation other than the 11 year cycle, an understanding of it could certainly influence the design of future space power systems.

Solar emissions related to the 11-year solar activity cycle cause major changes in Earth's magnetic field, plasmasphere, ionosphere, and atmosphere. In turn, there are induced effects at Earth's surface with significant socioeconomic consequences for energy technologies. Examples, some of which are illustrated in Figure 1, include:

1. Interruptions in long distance power transmission due to geomagnetic induction
2. Interference on oil and gas pipeline monitors
3. Severe interference with magnetometer signals obtained for geophysical exploration for minerals and petroleum
4. Modulation of both climate and weather.

Man-produced factors that are known to affect the natural solar-terrestrial system include:

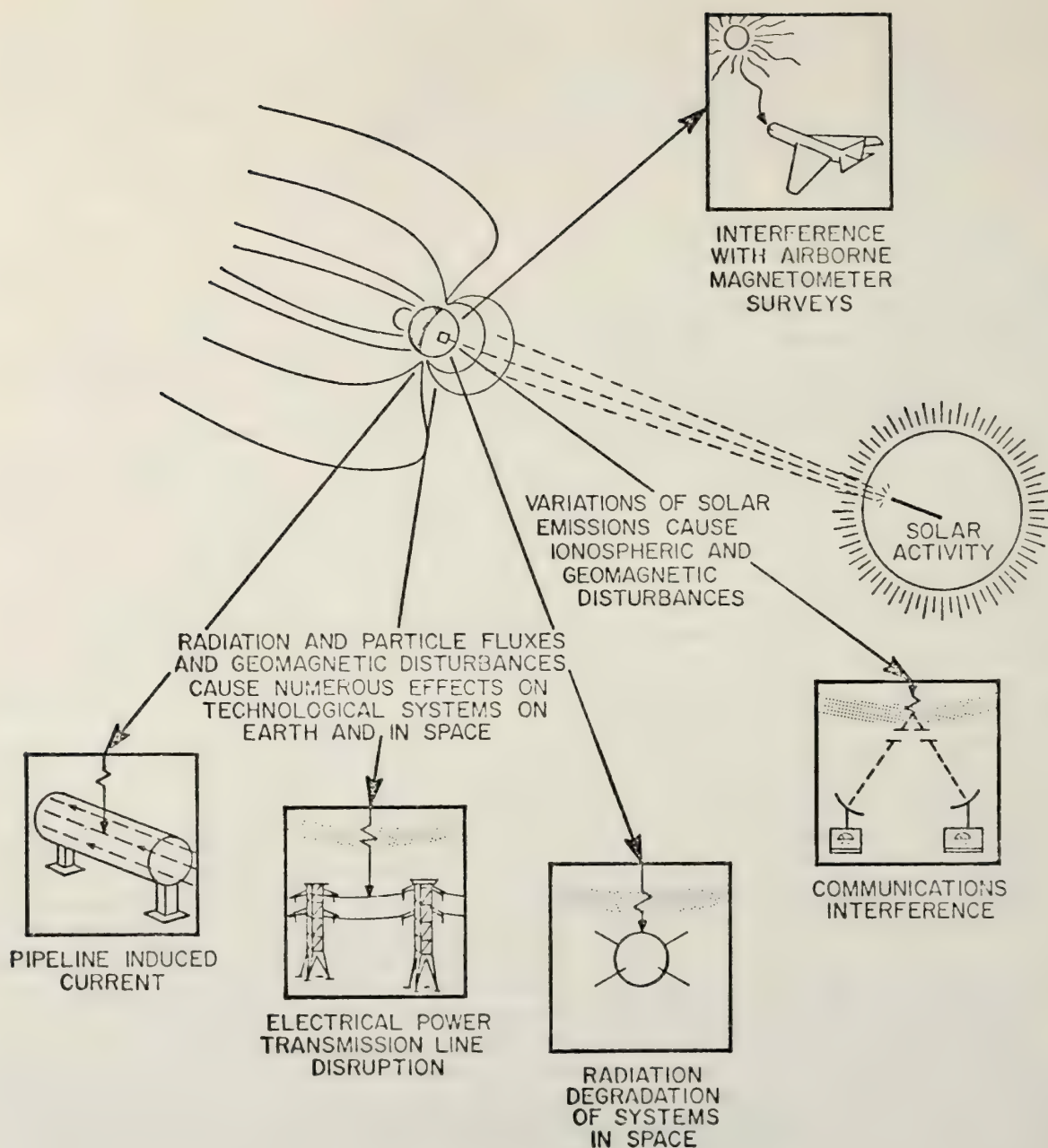


Figure 1. An illustration of how solar activity causes variations in Earth's outer environment, which in turn cause effects near Earth's surface.



1. Harmonics of 60 Hz and 50 Hz power systems may be a significant cause of electron losses from the magnetosphere into the ionosphere.
2. Increasing levels of carbon dioxide from fossil fuel consumption can cause a warming of the atmosphere and changes in climate.
3. Injections of various chemical and/or radioactive species into the atmosphere from energy technologies and other human activities may alter not only the total radiation balance that controls climate

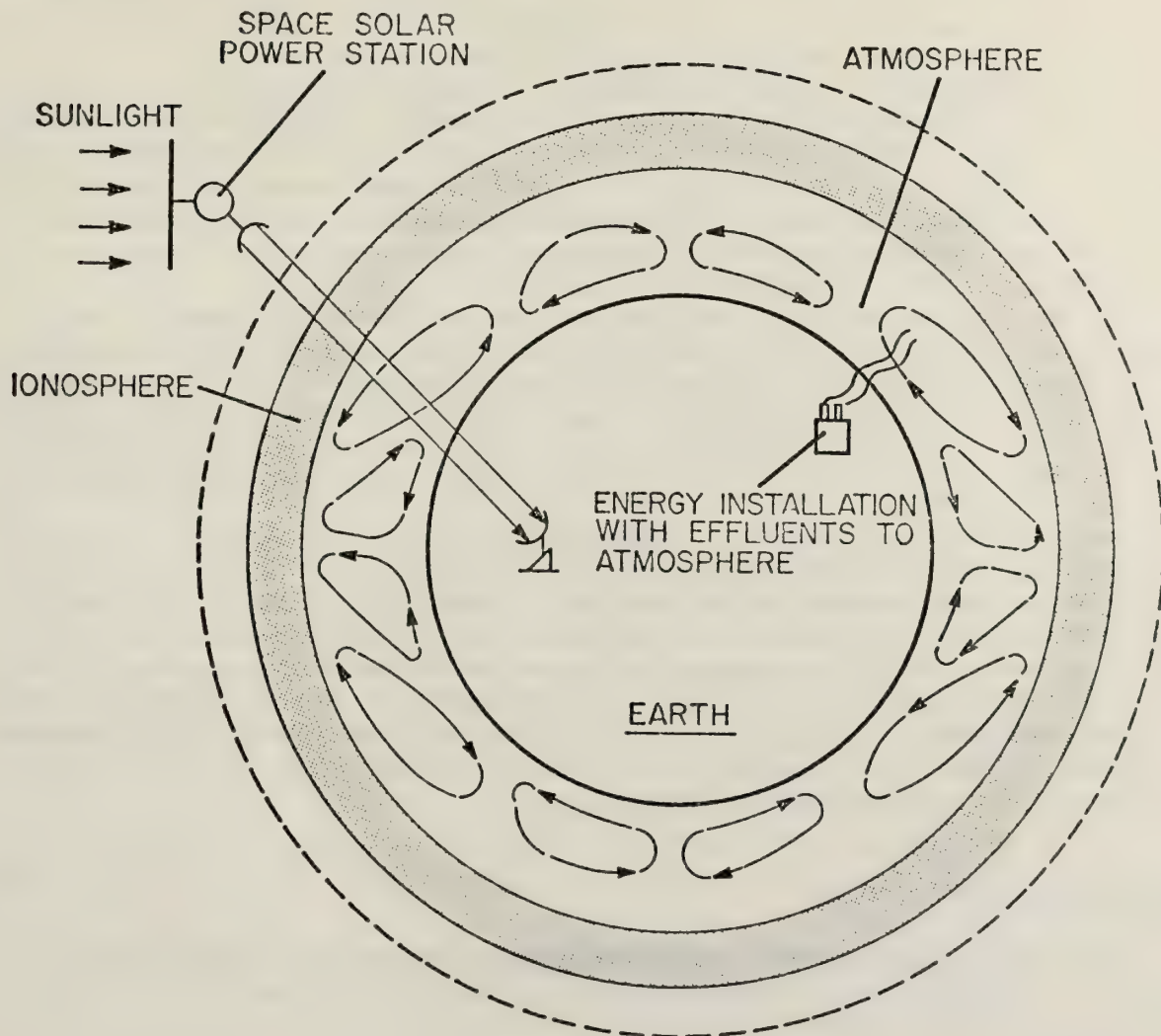


Figure 2. Earth's atmosphere and ionosphere (not to scale) to illustrate the potential global scale of energy-related activities. Large-scale circulation of the atmosphere and ionosphere transport can couple localized processes to regional or global scale effects.

directly but also the radiation balance in non-visible spectral regions, such as the UV, where there are human medical consequences. In addition, the atmospheric electrical state may be altered with possible meteorological consequences.

4. Space power systems have the potential to significantly modify the ionosphere locally; globally when used for long periods and large numbers.

Predictions are already in routine use for magnetometer surveys and pipeline monitoring. Past experiences with solar-terrestrial-caused power line interruptions and induced currents in large electrical generators have provided a basis for design of new systems. In planning the lifetimes of these systems, however, it will still be useful to have long-term predictions of major solar activity and the integrated amount of activity.

It is the area of solar modulation of climate that represents the greatest need for long-term predictions and understanding. We must first find out the degree of modulation caused by solar variability, of whatever nature, and whether it is on a global or regional scale. And the processes by which the influence is transmitted down to the troposphere must be understood before the predictions can be made. This is why we emphasize the basic research aspect of solar-terrestrial processes.

Figure 2 indicates in a very general way the global nature of the atmosphere and ionosphere. Although this figure is outmoded in its depiction of the general circulation of the atmosphere, it serves our purpose, which is to illustrate that events occurring locally can be diffused to global scales in some cases. Vertical and horizontal transport can occur in several ways, with the result that perturbations in the atmosphere or ionosphere do not always remain localized.

At present there is insufficient basic knowledge of climate, the atmosphere, magnetosphere, and sun as an interacting system to permit definitive conclusions on cause-effect relationships. Therefore, a need exists for basic scientific research to expand the knowledge base about this total system. However, before realistic assessments can be made about consequences to the natural environment from man's energy systems and vice versa, there will have to be much more emphasis on synthesis of knowledge about parts of the system into models of the whole system. Only then will we be able to realistically define specific prediction requirements.

#### Acknowledgment

This work was performed under the auspices of the U.S. Department of Energy.

#### REFERENCE

- Lanzerotti, L. J. (1978): Impacts of ionospheric/magnetospheric processes on terrestrial science and technology. Report edited by R. J. Lanzerotti and prepared for the Panel on Solar System Processes of the Study on Space Plasma Physics, Space Science Board, National Academy of Sciences.

### III. COMMUNICATIONS PREDICTIONS

Dr. A. P. Mitra, Working Group Chairman

#### A. IONOSPHERE-REFLECTED PROPAGATION

A WORKING GROUP REPORT prepared by B. M. Reddy, Subsection Leader

##### 1. INTRODUCTION

The scientific exploration of the Ionosphere and the study of its morphology took great strides in the 30 and 40s of this century as people realised the importance of the ionospheric media for long distance radio communications. However, it was the World War II that gave impetus to undertake systematic studies for predicting the ionospheric conditions to achieve reliable communication. Probably, the first such forecast from USA was given from the Carnegie Institute of Washington by A.H. Shapley in 1946. An excellent review of the historical forecasting methods was given by Meek (1970). The terms of reference of this working group are not concerned so much with the physical processes controlling the morphology and dynamics of the ionosphere (which are reported separately by working groups dealing with High latitude E- and F- region ionospheric predictions, Midlatitude and equatorial E- and F- region ionospheric predictions and D-region ionospheric predictions), but are restricted to discussions on the predictability of those ionospheric parameters relevant to Ionosphere-reflected communications and on their optimum utilisation. Several excellent original articles and review papers have been published from time to time dealing with the long-term and short-term forecasting of ionospheric parameters and with the radio systems and modelling needs for ionospheric communications (Bradley, 1975; Cook and McCue, 1975; Bradley, 1979a,b; Rush, 1976; Rush and Edwards, 1976). This report does not purport to give a review on the topic; only the most outstanding problems in view of the working group are discussed with the aim of emphasizing the future course of work in this area.

##### 2. HOW IMPORTANT IS IONOSPHERE-REFLECTED COMMUNICATION IN THE PRESENT DAY CONTEXT?

In spite of the growing use of satellites, microwave links etc. for communications, the importance of ionospheric communication continues to grow for a variety of reasons. The discussions in the various working groups led to the conclusion that HF communication is very much in the business and the number of users have considerably increased in the recent years. Undoubtedly, HF propagation via ionosphere will remain as one of the most important modes of long distance communication for a long time to come, especially for point-to-point communication. This is more so at low latitudes due to the following reasons:



- a. The middle and high latitudes are extremely vulnerable to the particle emissions from the sun, which play havoc with HF communications. But the low latitude ionosphere is safely protected from these particle invasions by the terrestrial magnetic field. Because of this reason, the major outages in short-wave communications at low latitudes will be limited to very short periods following solar flares (SIDs).
- b. Yet another factor which is a sequel to the solar event, is the ionospheric storm phenomena which are observable for a considerably longer time. These ionospheric storms usually produce large depressions in foF2 at middle and high latitudes causing enormous changes in the maximum usable frequencies. However, at low latitudes the changes in foF2 due to magnetic storms are comparatively much smaller. In addition, the changes in MUF at low latitudes due to the changes in foF2 and hmF2 are compensatory and so the disruption in radio communication due to storms is very rare.
- c. The foF2 values being much higher at low latitudes compared to high latitudes, the maximum usable frequency will also be higher. This facilitates a larger band of frequencies available for HF communications.
- d. The latitude belt encompasses a large number of developing countries, which can hardly afford very powerful Medium Wave transmitters for broadcast purposes and much less the expensive tropolinks for point-to-point communication systems. A sample calculation shows that a trans-receiving system in HF and for a distance of 2000 km can be operated with about 95% reliability at low latitudes with a cost of just 1% of that required for a microwave link of 99% reliability. Of course, the frequency channels and the bandwidth available will be far inferior in HF bands, but wherever the requirements are not stringent HF communications will be continued.

### 3. REQUIRED INPUTS FOR RELIABLE COMMUNICATIONS

- a. long- and short-term solar activity predictions
- b. prediction of foF2 and M(3000)F2
- c. prediction of foF1 and foE
- d. prediction of Es characteristics
- e. prediction of D-region behaviour
- f. prediction and summation of path probabilities and spread F

- g. prediction of radio noise - atmospheric, galactic and man-made
- h. prediction of magnetic activity, disturbances, SIDs and PC/ events

Several second order requirements can be added to the above as follows: prediction of antenna-terrain effects, time delay and inter-pulse interference, frequency-selective fading, phase stability of the ionospheric medium and polarization coupling loss.

#### 4. SOLAR ACTIVITY PREDICTION

While there has been a considerable amount of work in recent years with some degree of success on the short-term solar activity forecasting, the record has been near dismal in long-range forecasting. The status of knowledge in long-term forecasting of solar activity has remained where it was two or three decades ago; we are still left with either sunspot number or 10.7 cm flux observations in the past to be extrapolated into future by statistical techniques. This is one of the serious limitations in predicting the long-term median characteristics of the ionosphere. There was apparently no new input in this area in the workshop except description of old techniques in a number of prediction group review papers. In view of this, it is time that the ionospheric communication people develop techniques of predicting foF2 simply from the foF2 cycles during the last 45 years, short-circuiting the solar cycle forecast. The working group felt that the foF2 data series is still too short to arrive at meaningful statistical interpolation; but it should be tried by some groups and compared with solar predictions.

However, on the subject of short-term forecasting of solar flares, an overwhelmingly large number of papers were received for the workshop, reflecting the trend of work in this area. Several features like structure and strength of the magnetic fields of active regions, flare-produced changes in longitudinal field gradients, positions of flaring regions relative to  $H_1$  field, plage brightness in  $H_K$ , evolution of moustaches, surges and sunspot group rotation have been studied by the Russian group (Severny et al, 1979; Mogilevsky, 1979a,b). Soft x-ray emissions from the non-flaring sun as a flare precursor has been suggested by the group at NRL (Horan et al, 1979; Teuber et al, 1979) including a method of predicting the soft x-ray flux itself.

#### 5. PREDICTION OF IONOSPHERE-REFLECTED COMMUNICATION PARAMETERS

##### 5.1. Long-term Predictions

The existing techniques predict foF2, M(3000)F2, foE, D-region absor-



ption, foEs and fbEs wherever appropriate. Some elaborate techniques even use the ray-tracing program. Once again in this area no real advances were made in the last 15 years or so. In this context, the question often raised is whether extensive morphological measurements are more profitable from communication point of view than intensive study at a very few locations. Examples are the routine ionosonde stations versus rockets and facilities such as incoherent scatter. From the practical communicator's point of view, it should be emphasised that what is going to happen is much more important than why it happened. One may question with apparent justification, the utility of conducting more morphological studies even after more than four decades of such observations. But then, it should not be forgotten that the sun may change its course any time and may exhibit periodicities unknown so far. The reaction of the ionosphere to such changing solar conditions cannot be studied unless a continuous and balanced chain of observations is maintained. At one extreme, people are even wondering whether ionospheric predictions are necessary at all in this age of sophisticated electronics and systems control. For example, it is possible to have a HF link where the system can automatically and continuously search for the best operating frequency and tune itself for this optimum condition. Since the response time of such equipment can be much shorter than the response time of the ionosphere, it is possible to have such an adaptive communication system. The only problem in this naturally is how does anybody control the allocation of frequencies in the limited available spectrum? In fact the current status in many countries is that these frequency allocation organizations take anywhere from six months to two years to clear a frequency either for broadcast or for point-to-point communication purposes. So, the real need to be emphasized is to improve our long-term ionospheric prediction system. Improving a prediction technique not only needs long series of routine data but also observations from a few elaborate systems such as sophisticated ionosondes and incoherent scatter radars. The dynamic response of the ionosphere derived from such data should be combined with routine data to improve the prediction techniques over the current massive-data-oriented statistical methods.

There are fairly a large-number of groups around the world who produce long-term ionospheric predictions, basically using in some way or other the world maps of basic ionospheric parameters available in numerical form (CCIR 340). Some of the groups issuing independent predictions are from Japan (Maeda, 1979), Australia (Turner and Wilkinson, 1979), India (Reddy et al, 1979), FRG (Damboldt, 1979), Brazil (Picquenard, 1979), Poland (Stasiewicz et al, 1979; Klas and Stasiewicz 1979), USSR, UK, France and from several groups in USA (Thompson and Secan, 1979; Hatfield, 1979 and also ITS-78). Predictions one month in advance are known to enjoy better credibility. One such method is the MR method (Murthy et al, 1979) where a monthly ratio is used to multiply the preceding month's medians. The general observation that can be made is that while the prediction of D region,



foE and foF1 is satisfactory for long-term planning, the prediction of Es characteristics and foF2 is far inadequate.

## 5.2. Short-term Forecasting

A considerable amount of progress has been achieved in this area during last 10 years. This is mainly because of advances in short-term solar forecasting, availability of satellites monitoring X-rays, EUV and particle fluxes, and better understanding of the transient response of the ionosphere to events. Routine short-term forecasts are issued by Japan (Marubashi et al, 1979; Mori et al, 1979), Australia (Mc Namara, 1979), India, USSR (Avidiushin et al, 1979; Zevakina and Kiseleva, 1979), France (Lassudrie - Duchesne, 1979), FRG (Damboldt, 1979), U.K. and by several groups in USA.

In most cases, the short-term prediction is presented as a correction to the monthly predicted median values. The method used in Europe essentially consists of an extrapolation of the immediate past data with a correction based on geomagnetic and solar data. The IUWDS messages and presto messages from Boulder are used with great advantage. The data flow from local stations is in near-real-time in Europe and USA and the predictions consequently are found to be reliable.

## 6. PREDICTION OF MAGNETIC ACTIVITY, STORMS, SIDs, PCA EVENTS etc.

The prediction of a solar event responsible for SIDs is a topic for the working group dealing with short-term solar activity predictions. The response of the ionosphere to a predicted event or active center depends on the actual spectrum of radiations and particles likely to be emitted. The low latitudes are not subject to many of these disturbances except the SIDs. The departures in HF communication parameters due to magnetic storms are usually less than the day-to-day variability. Hence these predictions are of little relevance at low latitudes as far as ionospheric communications are concerned. However, all these predictions are essential at middle and high latitudes while particle predictions are a must at high and polar latitudes. A number of papers have dealt with ionospheric storm morphology and possible theoretical models to reproduce the observed F-region departures (for example, Matsushita, 1959; Matuura, 1972). The major problem in developing a prediction technique has been that it is found impossible to assign a certain change in foF2 or hmF2 for a given change in any of the common magnetic indices such as  $K_p$ . The F-region changes are complex dependents upon the phase of the storm, commencement type, universal time, local time, latitude, season and solar activity. The existing models are grossly inadequate and often misleading. Limitations in generalising any

prediction procedure can be appreciated from the attempt by Kuleshova et al (1979) to make foF2 forecasts if the time of main phase on-set is predicted.

With regard to solar flare particle predictions, there are essentially three warning modes of different time scales. The first one, that is the longterm mode, predicts the probability of flare occurrence during the entire transit of an active region during several solar rotations. The next mode is the prediction of particle fluxes a few hours to a few days in advance of the flare. A lumped parameter like the composite flare index of Dodson is found useful. The third mode, known as the short term reaction mode, involves the warning of the particle spectrum after the flare has occurred. Probably the only such warning that is currently operational is the PPS-76 system (Smart and Shea, 1979) that uses several real time data inputs. The technique is to generate a computerized time intensity profile of the solar proton intensity expected after the occurrence of a solar flare. The data used include H-alpha flare reports, radio flux measurements at a number of frequencies, soft x-ray, solar wind data and real-time particle data from satellites.

## 7. COMMUNICATION BY E<sub>s</sub> LAYER

This has mainly two aspects in communications. On the positive side, in certain areas like in the equatorial zone where it exists most of the time, it will be advisable to predict and use it for communications over large distances with marginal powers. On the other hand, this causes interference in FM systems in the VHF band because ranges as large as 2000 km are possible. The prediction method presented by Giraldez (1979) is comprehensive and shows dependence of the blanketing frequency (fbEs) on solar activity, solar zenith angle and latitude. Some predictions of E<sub>s</sub> are available for India (Saksena, 1975), South American Sector (Giraldez, 1979), USSR (Kerblay and Nosova, 1979) and North America. However no practical program exists as of now that can be fed to a computer along with H.F. skywave field intensity program on a global scale.

## 8. PREDICTION OF LF AND VLF PROPAGATION CHARACTERISTICS AND LUF VALUES

The D-region of the ionosphere enters the communication systems as the main absorber of MF and HF waves and as the reflector of LF and VLF waves. The D region reflects VLF and LF waves from heights of 60 to 75 km during day and from 80 to 90 km during night. In view of the importance of these frequencies in communication, Navigation, Time and Frequency transmissions and Loran C, it will be necessary to predict the wave guide and ground wave propagation probabilities for planning power requirements



and for estimating interference potential. Some very useful statistics of the disturbed D-region vis-a-vis radio communication are discussed by Larsen (1979). While predicting D-region characteristics it may be useful to use such solar indices that reflect x-ray contribution. In addition, this region is sensitive to mesospheric meteorology for which satellite radiances and D-region electron density profiles should be studied to develop predictability of the region. The empirical relationship given by Oyinloye (1979) for predicting both spatial and temporal variations of absorption (and hence LUF) take into account the 1-3 Å x-ray flux, the solar zenith angle and the magnetic dip of the station.

#### 9. COMMUNICATION PROBLEMS DUE TO DAY-TO-DAY VARIABILITY OF THE IONOSPHERE

The F2-region parameters of the ionosphere undergo large random daily variations which are apparently unrelated to any specific solar or geophysical event. While prediction of the median values of these parameters for any particular month can be made with some degree of accuracy, it has been impossible as of now to predict their day-to-day changes. Some studies have been made earlier for mid-latitudes (King and Slater, 1973; Rush et al, 1974) and a simple technique to index this variability at low latitudes has also been suggested (Aggarwal et al, 1979). In general, the variability in foF2 is large during night time and if this is compounded with the low night time values of foF2 the band available for H.F. communication is severely restricted. Enhanced foF2 variability, in addition to equatorial latitudes, has also been reported at Dip angles of 30-35°, especially during low solar activity (Potapova, 1966). Though it is difficult at this stage to attempt accurate prediction of this variability, it should be possible to indicate the geographical and seasonal extents when such variations may seriously deteriorate HF communications by giving monthly quality figures for specific geographic zones.

#### 10. COMMUNICATION PROBLEMS DUE TO LARGE TEMPORAL, VERTICAL AND SPATIAL GRADIENTS IN THE IONOSPHERE

The dynamic situation in the E and F regions during dawn and dusk hours and the large horizontal gradients in the F2 electron density in the equatorial anomaly region and the mid latitude trough region (at night) pose serious problems in predicting long distance communication parameters (Lakshmi et al, 1979). In case of long distance circuits in the East-West direction involving multi-hop F-region propagation, the problem of rapidly changing MUFs at the dawn location will extend to a large number of hours because the different F-region reflection points will fall in these transient locations at different periods. The large spatial gradients anomalously



cause two different MUF values between any two points along a meridian in the to-and-fro links if the point of reflection falls in a region of large horizontal gradients. This interesting situation results in a particular frequency being available in one direction, but not in the opposite direction. Prediction programs in future should include these gradients at least in the equatorial anomaly regions and in the mid latitude trough regions. Yet another serious problem is the non availability of data over oceans and over poorly-inhabited land areas. The present approach is to extrapolate the measured values at some discrete locations to those inaccessible areas by employing numerical mapping techniques such as the one developed by Jones and Gallet (1962). The inherent deficiency in this method can be appreciated by studying the foF2 correlation coefficients as a function of station separation distance (Rush and Miller, 1973; Rush, 1976). King and Samuel (1975) studied the accuracies of  $\Delta$  foF2 and hmF2 maps derived from CCIR models including comparisons with airglow data from satellites and found that most serious inaccuracies are found over South Pacific, South Atlantic and over the Indian ocean. It will be necessary to pool all the available topside sounder data and the total electron content data and go through a massive exercise to prepare world maps of F2-region parameters with better resolution.

Usually the MUFs are estimated from the predicted critical frequency of the layer and MUF factor which essentially contains the height information. However, the received signal intensities and the MUF values do depend to some extent on the electron density scale heights, especially in the F2-layer. Unfortunately very little is known about the global morphology of these vertical gradients to justify a prediction procedure using comprehensive ray-tracing programmes. Another hesitation in adopting such sophisticated prediction procedure is the day-to-day variability in F region parameters that might off-set any improvement achieved by such complicated techniques.

## 11. NOISE

In any communication system, the noise is the ultimate limiting factor to acceptable receiver performance. Excluding the internal noise of the system itself, the major contributions to the external noise come from the atmospheric noise essentially from thunderstorms, manmade noise including other unwanted radio transmissions and galactic noise. At frequencies used in ionosphere-reflected communications, the receiving system noise and the galactic noise are relatively unimportant. All the so called atmospheric noise predictions such as the popular CCIR 322 and the earlier documents such as NBS circular 462 and NBS circular 557 are compilations of the noise data collected at various locations and extrapolated into the future. CCIR 322, in addition describes the fine structure of the noise in statistical terms and also presents some methods to use the predictions in the

solution of operational problems. Lucas and Harper (1965) have also produced empirical representations by numerical methods and have generated separate coefficients for such seasonal time-period. The atmospheric noise being the biggest source of noise, especially in the tropical zone, many more measurements are needed for a statistically significant prediction technique to be developed. One such effort has been published by Indian workers (Saksena and Ghosh, 1978).

Note: This brief report reflects the contribution of a large number of participants during the workshop and contributions by correspondence before the workshop. Particular mention should be made of the inputs given by the working group members Dr. A.P. Mitra, Dr. D.R. Lakshmi, Mrs. S. Aggarwal, Mrs. Hatfield, Dr. Damboldt, Dr. Duchesne, Dr. Mangal Sain and Capt. Tascione.

## 12. RECOMMENDATIONS

1. The data base from ionosonde chain is the most important for ionosphere-reflected communications. This chain should be augmented both qualitatively and quantitatively. However, to evolve new prediction techniques it is necessary to conduct intensive studies at a few locations by such techniques as incoherent scatter and more sophisticated ionosondes.
2. Realizing that a majority of users do not have the capability to change the terminal parameters of their communication systems, the single most important goal in future work should be improvement in predicting long-term ionospheric parameters. This naturally calls for active work on long-term solar activity forecasting.
3. Ionosonde data and TEC data should be simultaneously studied to predict deterioration due to magnetic storms and day-to-day variability. This study will also yield estimates of slab thickness and electron content between E and F region peaks which are important in arriving at height factors and in ray-tracing prediction techniques. Recently obtained satellite radiance measurements should be combined with other D region measurements to improve upon the predictability of lower ionosphere.
4. The day-to-day variability of F region is probably the worst degrading factor in short-term forecasting. Global morphology of this variability should be evolved to give quality indices that can warn the communicators in advance.
5. Multipath fading and time delay in different modes should be quantified to be included in path loss calculations to preclude deterioration in digital communication.



6. It is necessary to predict and identify a few individual solar indices that are likely to control ionospheric parameters relevant to different regions of the ionosphere. (For example, an x-ray dominated flare might affect E-layer communication while a EUV dominated flare might affect F-region communication)
7. Efforts should be focussed on neutral atmospheric studies to derive predictable relationships in case of magnetic storms-induced variations and the day-to-day variability of the ionosphere. This is also extremely important for LF and VLF communication reflected by D region.
8. The attempts to arrive at high latitude ionospheric index using real time x-ray, solar wind, solar radio flux and solar particle data should be encouraged.
9. The spatial gradients associated with equatorial anomaly or mid-latitude trough and the temporal gradients associated with dawn and dusk phenomena cause anomalous communication problems. The currently available predictions should include predictions about these phenomena.
10. It is suggested that the working groups remain active and work by correspondence for the next one or two years to look back at the recommendations and to bring them to the attention of appropriate international bodies.

## REFERENCES

- Aggarwal, S., D.R. Lakshmi and B.M. Reddy (1979) : A simplified indexing of F-region geophysical noise at low latitudes. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Avidiushin, S.I., A.D. Danilov, A.B. Malishev, G.N. Novikova and P.M. Svidskiy (1979) : Forecasting of the ionospheric and geomagnetic conditions in the forecasting centre of IAG. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Bradley, P.A. (1975) : Long-term h.f. propagation predictions for radio circuit planning. The Radio and Electronic Engineer, 45:31.
- Bradley, P.A. (1979a) : Propagation at medium and high frequencies: Practical radio systems and modelling needs. AGARD lecture series, 99:3-1.
- Bradley, P.A. (1979b) : Propagation at medium and high frequencies : long and short-term models. AGARD lecture series, 99:9-1.



- Cook, F.E. and C.G. McCue (1975) : Solar-terrestrial relations and short-term ionospheric forecasting. The Radio and Electronic Engineer, 45:11.
- Damboldt, Th. (1979) : Propagation predictions for H.F. range by the research institute of the Deutsche Bundespost. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Giraldez, A.E. (1979) : Day time sporadic-E blanketing frequency predictions. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Hatfield, V.E. (1979) : H.F. communications predictions 1978 (An Economical up-to-date computer code, AMBCOM). Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Horan, D.M., R.W. Kreplin and K.P. Dere (1979) : Soft X-ray emission from the non-flaring Sun as a precursor to flare activity. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Jones, W.B. and R.B. Gallet (1962) : Representation of diurnal and geographic variations of ionospheric data by numerical methods. ITU Telecommun. J., 29:129.
- Kerblay, T.S. and G.N. Nosova (1979) : Statistical prediction of  $E_s$  - layer parameters and echo-signal characteristics. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- King, J.W. and A.J. Slater (1973) : Errors in predicted values of foF2 and hmF2 compared with the observed day-to-day variability, ITU Telecommun. J., 40:766;
- King, J.W. and J.C. Samuel (1975) : Accuracy of the CCIR F2 layer model at low and middle latitudes. Electronics letters, 11:366.
- Klos, Z. and K.K. Stasiewicz (1979) : Estimation of solar flare influence on radio circuit transmission loss. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Kuleshova, V.P., E.V. Lavrova and L.N. Lyakhova (1979) : Forecasting of foF2-variation for ionospheric disturbances. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Lakshmi, D.R., S. Aggarwal, P.K. Pasricha and B.M. Reddy (1979) : H.F. communication problems at low latitudes due to steep spatial and temporal gradients. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.

- Larsen, T.R. (1979) : Effects of D-region ionization on radio wave propagation. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Lassudrie - Duchesne, P., A.M. Bourdila and H.Sizun (1979) : The French short-term radio propagation predictions in the Decameter band. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Lucas, D.L. and J.D. Herper (1965) : A numerical representation of CCIR report 322 high frequency (3-30 Mc/s) atmospheric radio noise data. National Bureau of Standards Tech. Note 318, U.S. Government Printing Office, Washington.
- Maeda, R. (1979) : Radio propagation prediction sciences in Japan. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Marubashi, K., Y. Miyamoto, T. Kidokoro and T. Ishii (1979) : Forecasts of geomagnetic activities and H.F. radio propagation conditions made at Hiraio/Japan. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Matsushita, S. (1959) : The study of ionospheric storms morphology. J. Geophys. Res., 64: 305.
- Matura. (1972) : Theoretical models of ionospheric storms. Space Sci. Reys., 13 : 124.
- Mc Namara, L. (1979) : The use of Ionospheric indices to make real- and near-real time forecasts of  $f_oF_2$  around Australia. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Meek, J.H. (1970) : Historical outline of forecasting methods. Ionospheric Forecasting. AGARD conference proceedings, No.49.
- Mogilevsky, E.I. (1979a) : Probabilistic short-term prediction of solar activity I. Prediction of the evolution of active regions. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Mogilevsky, E.I. (1979b) : Probabilistic short-term prediction of solar activity II. Prediction of flares in flare-active regions. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.

- Mori, S., S. Yasue, Y. Munakata and K. Nagashima (1979) : Influence of sector polarity of the interplanetary magnetic field from the cosmic ray north-south asymmetry. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Murthy, P.S.N., C.S. R. Rao and Mangal Sain (1979) : Prediction of foF2 by the monthly ratio (MR) method. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Oyinloye, J.O. (1979) : Prediction of radio wave absorption in the ionosphere. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Picquenard, Armel A.E. (1979) : A simplified computer method for long-term calculation of HF sky-wave circuits. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Potapova, N.T. (1966) : Planetary distribution of the variability of maximum electron density in the F2-layer. Geomagnetism and Aeronomy (Tr. of Russian), 6:800.
- Reddy, B.M., S. Aggarwal, D.R. Lakshmi, S. Sastry and A.P. Mitra (1979) : Long term solar activity and ionospheric prediction services rendered by National Physical Laboratory. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Rush, C.M. (1976) : An ionospheric observation network for use in short-term propagation predictions. Telecommunication Journal, 43:544.
- Rush, C.M. and D. Miller (1973) : Some aspects of the day-to-day variability of the equatorial anomaly: American and Japanese sectors. Radio Science, 7:1085.
- Rush, C.M., D. Miller and J. Gibbs (1974) : The relative daily variability of foF2 and hmF2 and their implications for HF radio propagation. Radio Science, 9:749.
- Rush, C.M. and W.R. Edwards, Jr (1976) : An automated mapping technique for representing the hourly behaviour of the ionosphere. Radio Science, 11:931.
- Saksena, R.C. (1978) : Prediction of occurrence of sporadic-E in the Indian sub-continent. Ind. J. Radio and Space Phys., 4: 318.



- Saksena, R.C. and B.B. Ghosh (1978) : Prediction of atmospheric radio noise field strengths for the Indian Subcontinent. Ind. J. Radio and Space Phys., 7: 205.
- Severny, A.B., N.N. Stepanyan and N.N. Steshenko (1979) : Short-term forecast of evolution and flare activity of active region in Soviet works. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Smart, D.F. and M.A. Shea (1979) : PPS76-A computerized 'Event' mode solar proton forecasting technique. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Stasiewicz, K., M. Maksymenko and M. Jakimiec (1979) : KAZIA - A computer system for solar-terrestrial data processing and objective prediction. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Teuber, D.L., E.J. Reichmann and R.M. Wilson (1979) : A prediction method for the soft X-ray of solar flares. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Thompson, R.L. and J.A. Secan (1979) : Geophysical forecasting at AFGWC. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Turner, J.F. and P.J. Wilkinson (1979) : A weekly ionospheric index. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.
- Zevakina, R.A. and N.V. Kiseleva (1979) : On the short-term prediction of the space-time distribution of the auroral absorption. Submitted for publication in Solar-terrestrial predictions proceedings, Boulder.

## B. TRANSIONOSPHERIC PROPAGATION PREDICTIONS

A WORKING GROUP REPORT PREPARED BY J. A. Klobuchar, Subsection Leader, and Subsection Members, Santimay Basu, Sunanda Basu, P. A. Bernhardt, K. Davies, D. E. Donatelli, E. J. Fremouw, J. M. Goodman, G. K. Hartmann, R. Leitingner, M. J. Mendillo, M. PoKempner, R. G. Rastogi, C. L. Rino, J. Secan, D. G. Singleton and H. Soicher.

### 1. INTRODUCTION

This report addresses the current status and future prospects of our capability to make transionospheric propagation predictions. While there are also tropospheric effects on earth-space propagation, we will consider here only the effects of the ionized media, which dominate for frequencies below 1 to 3 GHz, depending upon the state of the ionosphere and the elevation angle through the earth-space path. The primary concerns are the predictions of time delay of signal modulation (group path delay) and of radio wave scintillation. Progress in these areas is strongly tied to our knowledge of variable structures in the ionosphere ranging from the large scale (thousands of kilometers in horizontal extent) to the fine scale (kilometer size). There are still numerous gaps in our knowledge of ionospheric variability (Rishbeth and Kohl, 1976) and there is still uncertainty regarding the relative importance of various mechanisms responsible for the time histories observed in total electron content (TEC), proportional to signal group delay, and in irregularity formation. The extent to which these gaps in our knowledge are meaningful for a particular system depends upon the individual system requirements; for some, a knowledge of monthly average climatology may be sufficient, while for others, the day-to-day variability of the ionosphere must be taken into account. In this report we discuss our current capability to make both short- and long-term predictions. The data base upon which predictions are made is examined for its adequacy, and the prospects for prediction improvements by more theoretical studies as well as by increasing the available statistical data base are examined.

#### 1.1 Present Prediction Capability

The present state of near term (a few months in advance) prediction capability of TEC monthly mean conditions is probably within  $\pm 20$  percent for regions where a time history of TEC exists. A large part of the uncertainty in making monthly mean TEC predictions a few months in advance is due to the uncertainty of future solar activity. Even if monthly mean TEC values could be predicted with 100 percent accuracy the short term approximately 25 percent rms deviation from monthly mean values would still limit daily predictions. Much of the deviation from monthly mean values is the result of ionospheric effects of geomagnetic storms, and average statistical patterns of TEC deviations from quiet behavior have been constructed for several stations in an attempt to improve TEC predictability.

The influence of lunar tidal effects upon TEC is also being studied. At present the best and only major improvement over monthly TEC climatology predictions can be obtained by a real data observation not more than a few hours old taken where the TEC-time delay correction is required.

Scintillation models have attempted to statistically describe the available observations and have characterized the day-to-day variability in terms of a magnetic disturbance index such as  $K_p$ . Present attempts at constructing amplitude scintillation models are severely data base limited, especially in the equatorial region where the seasonal behavior of scintillation at different longitudes is still not fully known. Attempts to use available spread F scalings of ionograms in the equatorial region to generate a realistic amplitude scintillation model have met with no success because of the lack of distinction made in routinely scaling range spread and frequency spread from the ionograms. Scintillation observations correlate well in occurrence with range spread, but not with frequency spread. For the sub-auroral latitudes amplitude scintillation models exist, but no good polar latitude scintillation data base is yet available.

Phase scintillation is highly geometry dependent, and, while a picture of average irregularity size, shape and orientation in both the high latitude and the equatorial regions is beginning to emerge with data from the DNA-Wideband satellite, the available geographic, and particularly temporal coverage of data, is small and will likely remain so in the near future. The prospects of predicting day-to-day variability of occurrence of scintillation in the equatorial region seem particularly bleak at present.

## 1.2 Future Transionospheric Prediction Prospects

For TEC, at least a background morphology exists in some statistical form, mainly from northern midlatitude stations. Severe data gaps exist in the near-equatorial and high-latitude regions and in much of the southern hemisphere. For scintillation data gaps exist in the polar regions and in the Pacific and Asian equatorial regions. More data is needed to improve statistical models and more effort should be expended by theoretical modelers to take into account the morphology obtained from available TEC and scintillation experimental data. It is unlikely that major improvements in day-to-day transionospheric predictions will be made in the near term; therefore users who really require corrections significantly better than that available from average climatology, and who cannot use the mitigation techniques to be described in the next section, must use a real-time updating service where near real-time values of TEC, and of scintillation, can be used to improve the prediction. Theoretical efforts on transionospheric propagation problems should continue, but the major prospects for improved future predictions will likely come as a result of an enlarged data base rather than a major theoretical advance.

## 1.3 Potential Systems Mitigation Techniques

There are techniques that systems designers may use to mitigate the effects of the ionosphere on transionospheric radio wave propagation. For ionospheric time delay mitigation systems designers may utilize two modulated frequencies with the RF carriers sufficiently spaced to allow the



direct measurement of the differential modulation or group path delay. An example of such a system where the ionosphere is made transparent to the operational user is the NAVSTAR/Global Positioning System in which identical 10 MHz modulation is imposed on carrier frequencies of 1227 and 1575 MHz and the modulation delay is automatically subtracted from the measured time delay at the 1575 MHz primary navigation frequency. In this system ionospheric time delay errors of the order of 100 meters in range are reduced to the same order of magnitude as other system errors, namely a few meters. In the TRANSIT satellite navigation system in which range rate information only is used for navigation, the rate of change of TEC is automatically compensated, to first order, by the use of two carrier frequencies.

To mitigate the effects of radio wave scintillation various time delay modems have been constructed in which the message is sent in a redundant fashion with a time delay longer than the correlation time of the scintillation. Space diversity schemes have also been proposed.

While these techniques are available, their expense and the constraints they impose upon a system may preclude their use. Therefore, predictions of ionospheric variability are required to formulate perturbation levels for which systems must compensate or adapt. For systems requiring short-term prediction capability, a possible approach is one which utilizes the best morphological model of the ionosphere updated with operational measurements from spacecraft or networks of earth terminals that record group path delay and scintillation parameters on a near real time basis.

#### 1.4 Present and Projected Near Real Time User Prediction Requirements

Designers and users of advanced navigation and communications systems are sometimes unaware of, or reluctant to discuss, the potential ionospheric effects on their systems. Thus, it is difficult to determine the exact requirements for future transionospheric predictions. The trend is towards increased use of space-to-earth links for higher communications data rates, and for greater accuracy in navigation, as with NAVSTAR/GPS, for example. Phase coherent detection has been used to extend range sensitivity of ground-based radar, and phase coherent processing used for a satellite-borne radar, such as the Synthetic Aperture Radar (SAR) on the short-lived SEASAT satellite. For these and any advanced navigation, communication, or surveillance system, the potential effects of phase scintillation will have to be considered. Currently, support in the scintillation area is limited to defining spatial and temporal climatologies for use in mitigating its effects on planned new systems. There is a need for improved climatologies and the definition of simple "rules of thumb" for qualitative short range warnings of severe scintillation. In the area of time delay, such as existing satellite tracking radars and deep space probe tracking radars, current user needs are being met adequately during periods of low solar activity and quiet geomagnetic conditions, but more effort is required in modeling TEC response to high solar activity and geomagnetic disturbances.

In developing future prediction capability consideration must be given to the limitations and requirements of forecasters. For real time applications computer models should be fairly small, quick-running, and designed for easy forecaster-model interaction; input parameters, whether observed or predicted, must be readily available; output parameters must be

tailored so that the customer can use them directly. The challenge for the future is to meet these needs while developing models that accurately reflect median conditions at any given time and location. There is a "noise" level below which no median model can predict, but which can be reduced by adaptive techniques that update the model in real time. Therefore, any model should also be designed to be easily adaptable within the system.

## 2. TIME DELAY PREDICTIONS

### 2.1 Monthly Median Time Delay Predictions

The time delay of transionospheric radio signals is directly proportional to the total electron content along the propagation path. Present numerical models are able to predict monthly median TEC to within an rms deviation of 20-25 percent in daytime and 30-35 percent at night at northern mid-latitudes. Historically, models of foF2 and MUF were developed by the Boulder Laboratories of the Department of Commerce primarily for use in predicting the performance of sky-wave communication systems, and have been accepted and distributed by the International Radio Consultative Committee for this purpose. However, as they comprise the only comprehensive global ionospheric representation, they serve as the basis for a reference ionosphere, and are used to construct electron density profiles for three-dimensional ray-tracing and to estimate TEC. Models of the two most important ionospheric parameters, monthly median foF2 and M(3000)F2, are used extensively in these applications (Jones et al., 1969; Jones and Obitts, 1970). These models were developed from a relatively limited data base: 1954-1958 for foF2 and 1954 and 1958 for M(3000)F2. The solar cycle which peaked in 1958 was the highest ever observed with abnormally high magnetic activity. Thus, the foF2 and M(3000)F2 models developed using data from this cycle should be checked using data from a more average cycle, such as the one with a maximum in 1968-69.

A comparison between the predictions of monthly mean values using the model and the actual observations of the monthly mean critical frequency of the F2 region (foF2), using the observed 12-month smoothed mean Zurich sunspot number, ( $R_z$ ) in the model, was made for every other year of the 12-year period 1960-1972. Data for these comparisons were obtained from approximately 125 observing stations and were limited to vertical incidence sounder (v.i.) data. The root mean square error (rms) of the deviations calculated for each month of this period was typically 0.5 MHz at solar minimum and 0.8 MHz at solar maximum. There was a strong seasonal pattern, with the lowest rms errors observed during May through July, and the highest in the months of September through March. The relative deviation was typically 12-13 percent, with only 5 months of the entire 84-month period having an rms error greater than 20 percent. These statistics are more representative of the Northern Hemisphere mid-latitude v.i. stations, which comprised about 40 percent of the data. Much greater errors are observed at high latitudes, at the equator, and in the Southern Hemisphere at all latitudes (only 30 percent of the data). The seasonal pattern in the Southern Hemisphere is similar to that at Northern Hemisphere stations, with the largest errors observed in local winter.



No analysis has been made to assess the accuracy of the M(3000)F2 numerical representations. The errors could be equally as large as, if not larger than, those observed in the foF2 representations. Primarily, this would be due to the smaller number of observations, poorer quality of data and the inadequacy of a defined relationship with solar activity. An analysis similar to the one performed on the foF2 model should be applied to the predictions of M(3000)F2 to determine the possible range of errors in estimating  $h_{\max}$ , which is derived from M(3000)f2.

Predictions of TEC are derived from electron density profiles, which are constructed from the models of foF2 and M(3000)F2. They therefore include the errors inherent in both these models. Comparisons of median observed TEC with predicted values at mid-and low-latitude stations (Mulkern, 1976; Ryan, 1979) have shown an error of 10 percent is not unusual at mid latitudes, with a 20-30 percent error often occurring at low latitudes.

Prediction errors, particularly in the mid-latitude Northern Hemisphere, may be reduced by a more realistic measure of solar activity. The present use of sunspot number and 10.7 cm flux may be an inadequate representation, and the use of a more realistic parameter, e.g. an ionospheric index or effective sunspot number (see Lassudrie-Duchesne et al., 1979; Reddy et al., 1979; McNamara, 1979) should be tested.

In the past, evidence has shown that foF2 varies with solar cycle only within the  $R_z$  range of 10 to 15 for solar minimum to about 150 for solar maximum, and saturates beyond these levels. For this reason it is felt that EUV may be a better indicator of solar variation. It is not known whether TEC is similarly affected, but with the high level of solar activity in the present cycle, the opportunity may now exist to examine this. If calibrated EUV is obtained from the Space Shuttle Program, these measurements should be made simultaneously with measurements of foF2 or TEC for correlation studies. Also, since solar cycle 20 was very close to the mean of all solar cycles and relatively quiet magnetically, it is reasonable to assume that a mapping of foF2 and M(3000)F2 data from this cycle, particularly for the period after solar maximum, could provide a more accurate reference model.

From recent evidence it appears that the model representations for some Southern Hemisphere v.i. stations are poor regardless of solar activity. The possibility of using other sources of ionospheric information, e.g., satellite observations and theoretical models, to improve the representations in the large ocean areas, particularly in the Southern Hemisphere and the equatorial regions, should be explored.

The development of the present global model concentrated on a more accurate representation of the equatorial region, resulting in less adequate representation of the high latitude ionosphere. Therefore, the use of polar ionospheric models should be considered and tested (Elkins and Rush, 1973). Other possibilities for improving high latitude predictions should be investigated, including changes in the IMF and its effect on TEC as well as foF2 (Zevakina and Lavrova, 1979), and correlation between TEC and foF2 and local magnetograms.

## 2.2 Lunar Tidal Effects

Possibilities for reducing model prediction errors must include incorporating models for sources of variation in TEC other than solar effects. One of these is a model for lunar tidal variations (Bernhardt, 1976) which may comprise as much as 15 percent of the variation in TEC.



Lunar tidal oscillations are induced in the earth's atmosphere by the moon's gravitational field. They raise or lower the F region by direct action on the ionosphere via tidal winds; and by indirect action in creating dynamo electrostatic fields in the F region that couple into the F region along geomagnetic field lines producing ExB drifts. As the F region is moved to altitudes with different loss rates, its electron content (and peak concentration) is changed. Although there is no supporting evidence, it is possible that lunar tides can change the composition of the neutral atmosphere, which would change the F-region loss rate at a given altitude and produce TEC variations.

Since the lunar tide produces modulation of the existing ionosphere, lunar predictive models should be applied as multiplicative (as opposed to additive) corrections to existing models that only consider solar effects. Thus, a model for electron content, including the lunar effects, should be of the form

$$I(t) = S(t)\{1 + M(t)\}$$

where  $I(t)$  is electron content,  $S(t)$  is the solar diurnal variation of electron content, and  $M(t)$  is the lunar contribution. Initial verification of this formula has been demonstrated by Bernhardt (1978). Ideally,  $M(t)$  has a period of 12.42 hours, one-half of a lunar day, and is dependent on latitude, dip angle and season. Some characteristics of the function  $M(t)$  have been described by Rao and Stubenrauch (1967), Bernhardt et al. (1976), and Huang (1978).

### 2.3 Plasmaspheric Contribution to TEC

A contribution to TEC that requires further examination before being incorporated into an existing model is the plasmaspheric electron content. TEC, which is the integrated electron content along a path through the ionosphere, is primarily obtained through Faraday rotation measurements, which measure the content to a vertical height of approximately 2000 km. The electron content of the earth's plasmasphere, defined here as the region above the height ( $\sim 2000$  km to  $\sim 30,000$  km) where measureable VHF Faraday rotation occurs, extending to the plasmopause, has only recently been measured by transionospheric propagation techniques. Experimentally, the plasmaspheric content,  $N_p$ , has been determined by taking the difference between the total electron content,  $N_T$ , along the entire satellite to ground station path as measured by the group delay of 1 MHz modulation envelopes on RF carriers at 140 and 360 MHz transmitted from a radio beacon on the geostationary ATS-6 satellite, and the ionospheric electron content,  $N_F$ , inferred from the Faraday rotation along the same path. Since the ATS-6 satellite launch in May 1974, the behavior of  $N_p$  over the North American, European and Asian sectors has been reported by Soicher (1976a,b), Davies et al. (1976), Klobuchar et al. (1978), Poletti Iuizzi et al. (1977), Hajeb-Hosseini (1977) and Fleury and Corme (1978) for a period near the minimum of the solar sunspot cycle.

Generally, the ratio  $N_p/N_F$  varies from  $\sim 15$  percent during the day to  $\sim 50$  percent during the night. Some diurnal and seasonal variations are observed but in absolute value (as compared to  $N_F$ ) the changes are rather small.  $N_p$  increases with decreasing geomagnetic latitude marked by the ray path to the satellite.

To date, no data are available for the time period of increasing solar activity, and with the sporadic performance and planned shut off of the ATS-6 beacon experiment, the behavior of  $N_p$  during sunspot maximum may not be known. It is expected, however, that  $N_p$  will increase with solar activity but not to the same extent as  $N_F$ . During magnetic storms the pattern of  $N_p$  behavior shows a significant depletion with a gradual return to average values after a period lasting from three to eight days (Soicher, 1976c; Poletti-Liuzzi et al., 1977; Davies et al., 1976; Kersley and Klobuchar, private communication). In addition there is some evidence of a positive phase in  $N_p$  during the initial phase of a magnetic storm (Soicher, 1976c; Degenhardt et al., 1978).

In view of the small absolute value and relatively small diurnal variability of plasmaspheric content in comparison to the ionospheric content, especially during the day when the effects of TEC in operational satellite navigation consideration (i.e., excess propagation time delay) are largest, it seems that simple additive constant values to the ionospheric content of 15 percent of noontime  $N_F$  may account sufficiently for  $N_p$ . Predictions of  $N_p$  for quiet and magnetically active periods would thus be based on the same prediction techniques that apply to ionospheric TEC.

The above comments are based on relatively sparse data from several stations in the Northern Hemisphere at geomagnetic latitudes from  $30^\circ$  to  $60^\circ$  during solar minimum conditions. Additional data at diverse locations during maximum conditions are necessary in order to obtain global morphology of the plasmasphere under varying solar conditions. Prospects of obtaining additional  $N_p$  data for the next few years are poor as simultaneous group delay and Faraday data are required from the same satellite and ATS-6 was the only satellite with such a combination of appropriate beacon transmitters. The major continuing source of synoptic data on electron content will still be Faraday rotation measurements.

## 2.4 Geomagnetic Storm Effects on TEC

A major source of variation in TEC from monthly mean conditions is geomagnetic activity. Since theoretical modeling capabilities are not sufficiently advanced to predict TEC storm behavior, morphology-based prediction schemes represent the only viable method currently available. The major needs for the transionospheric propagation community pertain to times and places when daytime positive ionospheric storm effects ( $\Delta\text{TEC} > 0$ ) may be expected. In this respect, Mendillo and Klobuchar (1979) typified the state-of-the-art of TEC storm-associated prediction capabilities. Earlier, Mendillo (1973) described how one could predict whether or not a given storm (i.e., one in progress) would be expected to produce a positive phase at mid-latitudes.

Theoretical explanations for the positive and negative phase variations of TEC storm behavior generally deal with electro-dynamic processes for the TEC enhancements (see, for example, Anderson, 1976) and with changes in the neutral composition ( $O/N_2$ ) for the F-region depletions (see, for example, the recent work of Prölss and van Zahn, 1977).

The major uncertainty encountered in attempting to apply morphology-based TEC storm correction terms to monthly median behavior centers on the need to predict the storm commencement (SC) time, for both sudden storm commencements (SSCs) and gradual storm commencements (GSCs). The geophysical



parameters that, if provided via predictions or real-time measurements, would offer the greatest insight into storm effects, are neutral atmospheric parameters (winds and  $O/N_2$  compositions) and storm-time electric fields.

The world regions where TEC storm morphologies (and therefore prediction capabilities) are lacking are (1) low and equatorial latitude regions and (2) the entire southern hemisphere.

The main areas of F-region storm analysis that have led to inconclusive results are (1) gross averaging techniques which ignore the dual-phase and multi-day nature of storm effects and (2) correlations between ionospheric variations and various geomagnetic disturbance indices (e.g.,  $\Delta\text{TEC}$  vs.  $K_p$  or  $\Delta\text{Dst}$  relations). These schemes appear to be of little value for predicting storm effects.

## 2.5 Short Term Time Delay Predictions

Monthly mean time delay values, even with corrections for average magnetic storm induced changes in TEC, may not be sufficient for certain systems or for some classes of users of systems. They may instead require near real-time measurements of the local ionosphere, either of  $f_oF_2$  or of TEC, to update median models. These updating techniques include producing improved weekly or monthly mean values to serve as a basis for predicting values for the following several hours.

Studies of adaptive techniques that use real-time observations to reduce average monthly rms errors in predictions have been conducted using  $f_oF_2$  data (Wilkinson, 1979) and TEC data (Donatelli and Allen, 1978, 1979; Leitinger et al., 1978). The potential for reducing this rms error in TEC at a specific location is demonstrated by Figure 1, using data from Hamilton, Massachusetts, for January 1969. Figures 1a and 1b compare the results using an "ideal model," i.e., the observed mean (a) versus the predicted values (b). First note that the observed mean is lower than the prediction for the entire daytime period, while the rms deviation is comparable. The dashed lines indicate the rms residual error for the month using an observation after three hours to update the mean each day. The error growth rate is described by the curves originating at two-hour intervals with zero-error, and continuing in time, for three hours in the observed case (a) and twelve hours in the predicted case (b); the observation at the hour is used for updating at 15-minute intervals to describe the curves. The criteria for determining the maximum observation-to-update interval time is that the residual error when updating does not exceed the monthly residual error as represented by the standard deviation. A three-hour observation is useful throughout the daytime hours, unless the measurement was made prior to sunrise and used thereafter. The error introduced after sunrise is greater for case (b) than case (a), and the reduction in residual error is not as significant for case (b), which represents present model potential. As stated previously, existing models were developed from data obtained in the 1954-58 period, and may be biased toward both high solar and high geomagnetic activity. January 1969 was a relatively quiet month, magnetically. This may explain the up to 20 percent difference between observed and predicted daytime values, discussed by Ryan (1979).

The following conclusions can be stated from examining the error growth in time.



# HAMILTON JANUARY 1969

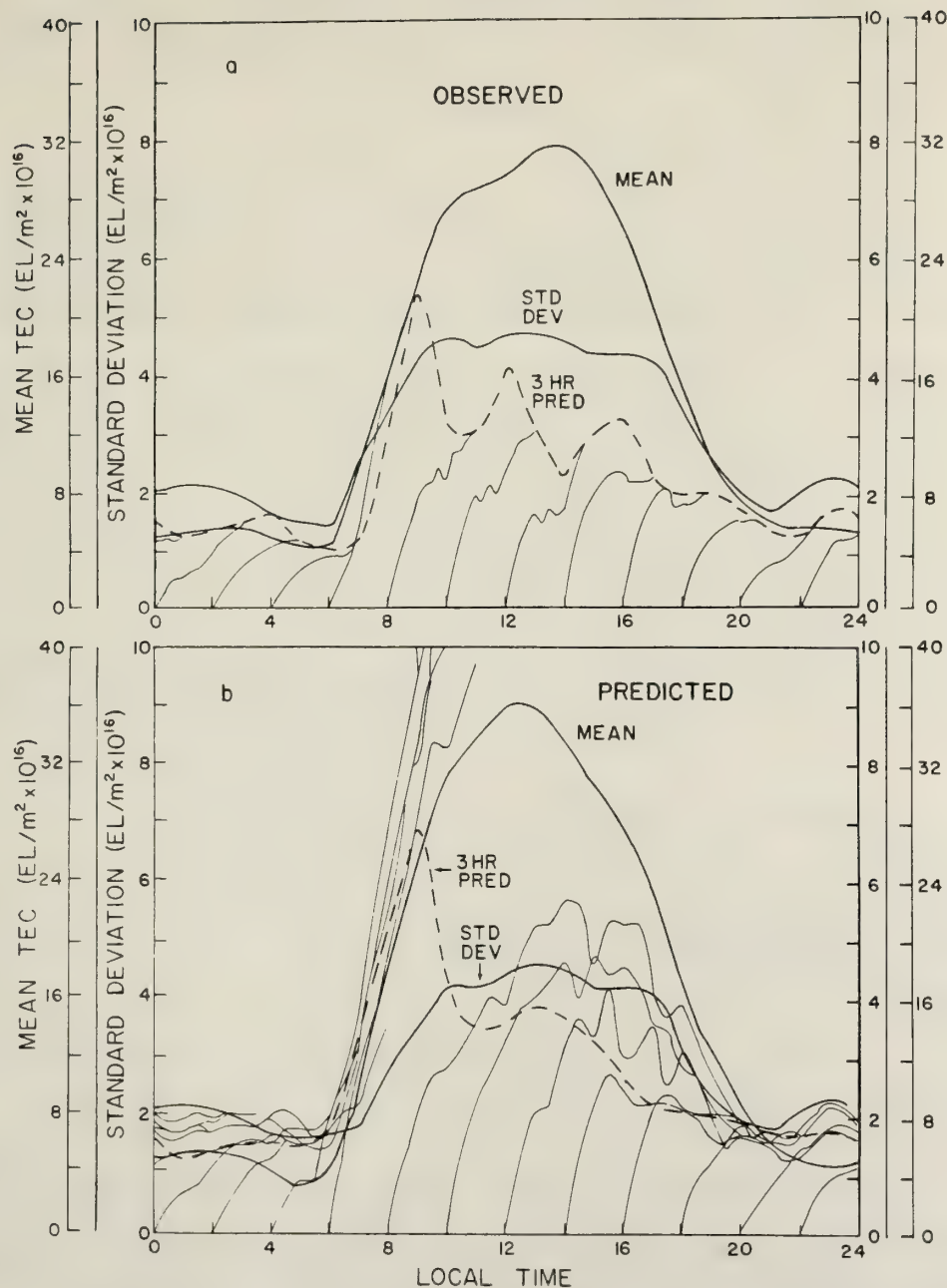


Figure 1. Observed mean (a) versus predicted values (b) of mean TEC.

1. Both observation and update should occur within the following intervals (projection of a prediction across terminators may introduce large errors):
  - a. post-sunrise to sunset
  - b. post-sunset to sunrise
2. For appreciable reduction in residual error (about 50 percent), the intervals should be less than the following:

- a. solar maximum
  - i. daytime, 3 hours
  - ii. nighttime, 1 hour
- b. solar minimum
  - i. daytime, 1 hour
  - ii. nighttime, 30 minutes

While the intervals for useful updating during solar minimum may be shorter to effect the same percentage reduction, it should be noted that the absolute error is much less to the extent that for the same month, the standard deviation at solar minimum may be less in absolute value of TEC units than the residual error at solar maximum using the updating intervals recommended in (2).

Spatial projection of observations have been discussed in correlation studies using foF2 (Rush, 1976) and TEC (Klobuchar and Johanson, 1977), and indicate a latitudinal correlation that is approximately one-half the longitudinal correlation distance. This must be considered with the time effects for any adaptive technique in which real-time observations are used for updating at a location other than that from which the observation was made. This has been examined using TEC data from mid-latitude stations (Donatelli and Allen, 1979; Allen et al., 1977). The increased error growth rate represents a superposition of spatial and temporal fluctuations. By reducing the space-time interval to an equivalent time interval, a useable updating interval may be estimated. A good approximation can be made with the following:

$$\frac{1}{15} \{(\Delta \text{LON})^2 + (2\Delta \text{LAT})^2\}^{\frac{1}{2}} + \left\{ \begin{array}{c} \Delta T \\ \text{or} \\ T - \Delta LT \end{array} \right\} = \Delta ET$$

where:

$\Delta \text{LON}$  = longitudinal separation, in degrees

$\Delta \text{LAT}$  = latitudinal separation, in degrees

$\Delta T$  = time interval between observation and update, in hours

$\Delta LT$  =  $\Delta \text{LON}/15$ , in hours

{ } refers to choice for east or west direction from observing to updating location;  $\Delta T$  if  $W \rightarrow E$ ,  $|\Delta T - \Delta LT|$  if  $E \rightarrow W$ .

The expected residual error can be estimated by setting  $\Delta ET = \Delta T_t$  where:

$\Delta T_t$  = time interval between observation and update at same location.

Then  $\Delta T_t$  is subject to the constraints in 1 and 2 above. Whether this holds for high and low latitude locations, remains to be tested. At present, adequate data are not available to make this study.

## 2.6 Theoretical versus Empirical Models of Transionospheric Propagation Predictions

Predictive models based upon both a theoretical and an empirical basis may be of more use than either a physical model or a morphological model alone. There is no particular reason why predictions should be made with either a purely physical or a purely morphological model. Hybrid models, which are part physical and part empirical, may be optimum. The advantages of

physical models are: (1) they enable assessment of the relative importance of ionospheric responses to isolated changes, e.g., neutral composition, ionizing flux, dynamics etc.; (2) they can provide predictions of parameters not readily observable, e.g., ion composition, plasma temperatures, etc.; and (3) they can be used to provide predictions of artificial (man-made) ionospheric changes; for example, depletions of the F-region plasma by rocket exhaust has been successfully predicted.

Physical models also suffer from a number of deficiencies: (1) all physical processes may not be included; (2) those processes that are included may not be treated adequately; (3) the numerical procedures available for the modeling may be inadequate and too time consuming; (4) inputs often cannot be determined reliably, e.g. solar ionizing flux; (5) the temporal and spatial variation of inputs such as neutral composition is only known on the average, and on a large scale; and (6) the dynamics of the inputs are unknown.

In contrast to physical models there exists a class of purely morphological (empirical) models. These models use measured ionospheric characteristics with firm statistical evidence in some regions and extrapolation of unknown accuracy to other regions. The disadvantages of empirical models arise from their non-theoretical nature: (1) they cannot be used to extrapolate to parameters not previously observed; (2) creation of these models requires a large geographic and temporal network of observation stations; (3) data recorded at these stations requires consistent calibration and reduction procedures; (4) if the models are keyed to some geophysical index, a method of predicting this index must be provided; and (5) they cannot be used to predict man-made effects on the ionosphere.

In summary, theoretical and empirical models should be treated as complements. The flexibility of the theoretical model is required to predict specific effects. For this reason, and for the hope of significant future improvements, further theoretical work should be encouraged. The empirical models are best suited for predictions of a general nature at present.

## 2.7 Improving Transionospheric Predictions

To improve present climatological models used for transionospheric propagation predictions, the primary need is for more and better data with greater spatial resolution. In addition, there is a need to measure parameters from the neutral atmosphere that may provide insight as to the reasons for the complexity in the spatial variability of TEC. An example of this complexity, for a limited latitude range is shown in Figure 2, from Hartmann et al. (1979). If an increased spatial density of observations is to be of widespread use, future ground based observations of satellite beacon transmissions for TEC and scintillation must have standard data formats, calibration techniques, data editing, processing and interpretation techniques. Increased data gathering creates problems that grow with the data growth rates, particularly for stations which have little experience handling large amounts of data. A discussion of problems associated with the data (information) explosion and its potentially negative consequences can be found in Hartmann (1978) and Hartmann (1979).

To fully realize the objective of obtaining new transionospheric data to improve morphological models, data gathering centers like the World Data Center A at Boulder, Colorado, must play a very active mediator role that



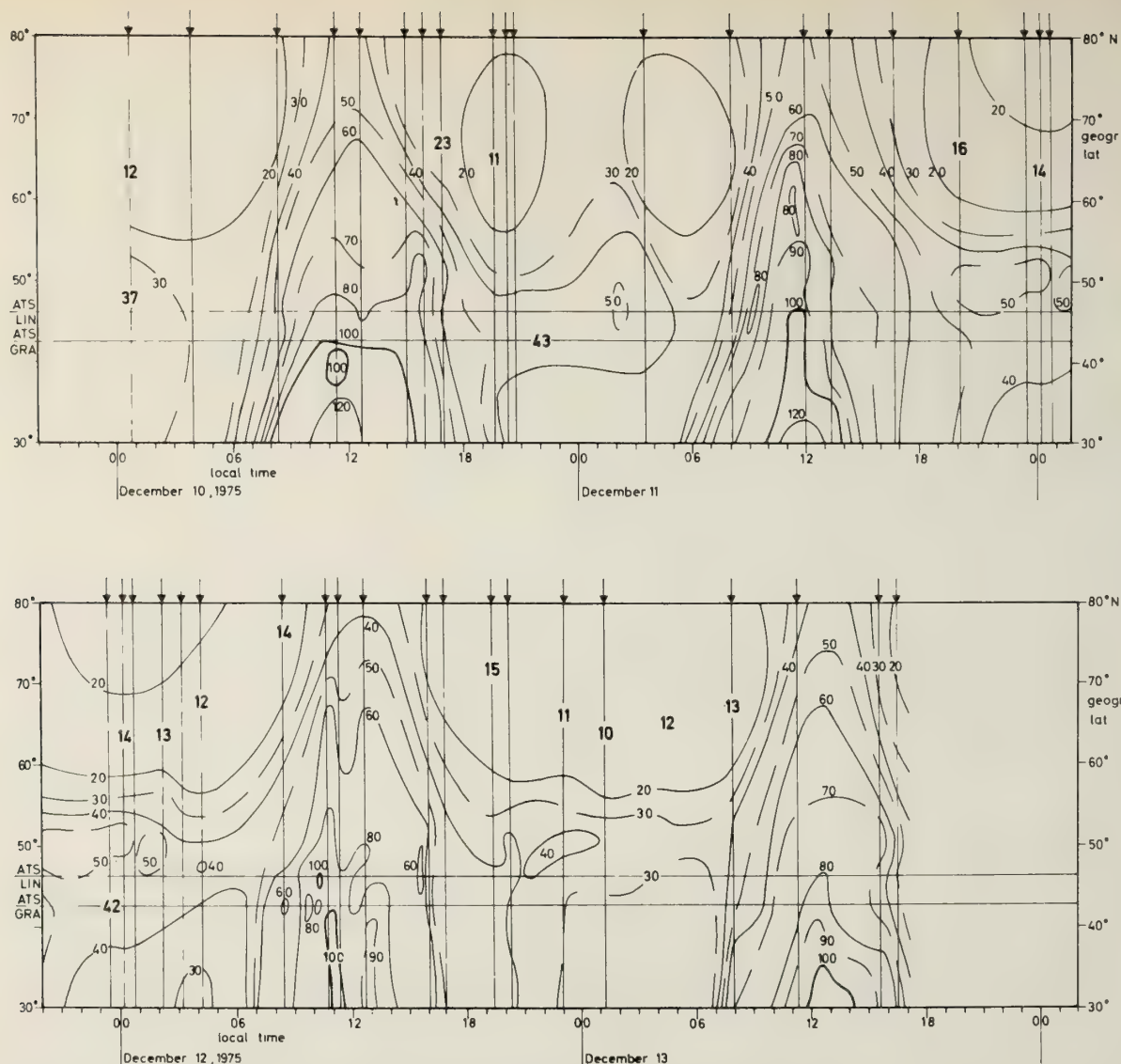


Figure 2. Contours of vertical ionospheric content from 30°N to 80°N geographic latitude in units of  $10^{15}$  el/m<sup>2</sup>. The maps are constructed by a combination of data from the ATS-6 synchronous satellite with data from the polar-orbiting Navy navigation satellite system satellites (after Davies et al., 1979).

will go beyond their present activities to insure that all accepted data conform to the standard requirements. In the last section of this report, specific recommendations are made for future data gathering to improve morphological transionospheric prediction models.

### 3. SCINTILLATION MODELING

#### 3.1 Background

Irregularities in the ionospheric electron density are known to cause potentially disruptive radio wave scintillation effects. Transionospheric

radio communications are primarily affected by amplitude scintillation (fading). The severity of the amplitude scintillation has been characterized by a number of scintillation indices (Bischoff and Chytil, 1969; Whitney et al., 1969) which can be interrelated (Whitney, 1974). However, the index

$$S_4 = \left\{ \frac{(I^2) - (\bar{I})^2}{(\bar{I})^2} \right\}^{\frac{1}{2}} \quad (1)$$

where  $I$  is the signal intensity (amplitude squared) has proven to be most useful for theoretical analyses and modeling.

The probability distribution of the amplitude scintillation is accurately reproduced by the Nakagami distribution (Nakagami, 1969) for a wide range of  $S_4$  values (Whitney et al., 1972; Fremouw et al., 1979). Moreover, the Nakagami distribution is completely specified in terms of  $S_4$ . Thus, if  $S_4$  is known, the fading depth can be computed. Indeed, if the fade coherence time,  $\tau_c$ , is long compared to the time used in transmitting one data bit or baud, the average bit error can be computed in terms of  $S_4$ . Coding or interleaving can be used to recover the performance loss, but it is very costly to implement in a slow fading environment.

As  $\tau_c$  decreases, the "coding gain" generally increases until  $\tau_c$  becomes comparable to the baud duration. At that point the channel fluctuations distort the waveform, and catastrophic failure ultimately results. For typical ionospheric  $\tau_c$  values and data rates as low as 75 bits per seconds, pulse distortion only occurs under conditions of strong scattering where  $S_4 \sim 1$  and the Rayleigh fading model is a good approximation to the scintillation structure.

The performance of various communication systems in a Rayleigh fading channel has been extensively analyzed. As a practical matter, subtle time structure changes have only a small effect, and only the coherence time  $\tau_c$  need be specified to characterize the channel (Whitney and Basu, 1977).

It has been tacitly assumed that the coherence bandwidth is large enough to support the modulation scheme that is being used. Generally this is a safe assumption. Indeed, most communication systems that utilize large bandwidths do so with frequency hopping. Loss of frequency coherence between hopped frequencies will actually improve performance in a well-designed system. Thus, a rough measure of coherence bandwidth is generally adequate for communication systems applications.

These results can be summarized as follows: Insofar as the performance of transionospheric communication systems are concerned, only  $S_4$  and  $\tau_c$ , together with a coarse measure of coherence bandwidth, need be specified. The complex signal (amplitude and phase) is only relevant under fast fading conditions, but then the Rayleigh model, which only depends on  $\tau_c$ , can be used.

There are, however, potential (and actual) new satellite and ground-based radar systems that attempt to maintain phase coherence over long time or spatial intervals. Moreover, angle of arrival fluctuations demand knowledge of phase. On a somewhat larger scale, group-delay corrections must be made if precision navigation systems are to achieve their full potential. Thus, there is a need for models that predict both the large-scale and small-scale ionosphere-induced phase variations as well as the corresponding amplitude scintillation structure.

Here, however, progress in propagation theory as well as the TRANSIT,



ATS-6, and WIDEBAND differential phase measurements have considerably simplified the magnitude of the modeling task. An important finding from that work can be summarized as follows: If the integrated electron content structure can be characterized over a scale size regime that encompasses the largest and smallest significant scale sizes, then all ionosphere-induced phase and amplitude perturbations can be accurately and efficiently modeled.

The underlying assumption here is that the ionosphere can be replaced by a single equivalent phase-changing screen at an appropriate equivalent height. Fully accommodating the integrated diffraction effects through the ionosphere adds considerable complexity to the problem, but it does not significantly change any predictions based on single phase screen computations (Bramley, 1977).

### 3.2 Presently Available Amplitude Scintillation Models

Since the integrated electron content structure at present cannot be adequately characterized over a scale size regime that encompasses the smallest significant scale sizes to the largest ones, then we must resort to morphological models using available empirical scintillation data. The Fremouw and Rino model (Fremouw and Rino, 1973), suitably expanded by the use of the magnetic modifier (Fremouw et al., 1977), does not adequately reproduce newly available scintillation observations. Small summaries of scintillation data useful for systems planners have appeared in several forms. In the field of signal statistics, a series of measurements have been made at high and equatorial latitudes (Whitney, 1974; Whitney and Basu, 1977; Yeh et al., 1975; Fremouw et al., 1978). These have not as yet been put together to form morphological models of signal statistics. However, sufficient examples have been produced to allow the amplitude spectra to serve as examples at various levels of scintillation activity at high and equatorial latitudes.

In the field of amplitude scintillation models, simple algorithms have been constructed to reproduce several years of data at 137 MHz at auroral latitudes (Aarons et al., 1978). The effect of magnetic activity, the solar flux forcing function, the seasonal variation, and the local time are inserted into the algorithm which allows the average value of 137 MHz scintillations to be determined for a narrow swath of invariant latitudes ( $60-64^\circ$ ). No good polar data base is available. In the equatorial region, another series of algorithms was developed to reproduce observations from Accra, Ghana, and from Huancayo, Peru. The former station showed sensitivity to magnetic variations, the longitude of the latter station yielded only small perturbations of scintillation activity with magnetic index. However, the algorithms now serve as only the reproduction of the diurnal pattern; other observations have not been merged into a coherent set of data or, in the case of spread-F, have not been put into terms of scintillation index.

### 3.3 Parameterization of Scintillation Models

To predict  $S_4$ , one first resorts to the weak-scatter theory. The first formula that was used for modeling was developed by Briggs and Parkin (1963). It was based on a gaussian spectral density function. When it was later found that a power-law model was more appropriate, Rufenach (1975) revised the Briggs-Parkin model. However, he retained the propagation angle depen-



dence appropriate to the gaussian spectral density function. A corrected version of Rufenach's formula was later developed by Costa and Kelly (1976).

More recently Rino (1979a,b) developed a model that accommodates the effects of the large outer-scale cutoff that is appropriate to the ionosphere for both phase and amplitude scintillation. Moreover, the model allows for a variable spectral index, whereas the Rufenach and Costa and Kelly models allowed for only a single value. In any case, the basic model parameters are the following:

$C_s, \nu$	The irregularity spectral density function has the form $C_s q^{-(2\nu+1)}$
$a, b, \delta$	Axial ratios along and transverse to the magnetic field and orientation angle of second axis (Rino and Fremouw, 1977)
$h, L$	Effective height and thickness of equivalent phase screen

If any outer scale wavenumber,  $q_0$ , is specified, the rms electron density is given by the formula

$$\langle \Delta N_e^2 \rangle = \frac{C_s q_0^{2-2\nu}}{8\pi^{3/2}} \Gamma(\nu - 1) / \Gamma(\nu + 1/2). \quad (2)$$

If the source and receiver locations and velocities and the frequency are specified,  $S_4$  can be computed. Moreover, simple bounds for  $\tau_c$  can be derived for both weak and strong scatter. Finally, the measured phase variance can be estimated for a specified measurement interval.

True predictive modeling hinges on our ability to determine the dependences of these parameters on latitude, longitude, time, day, magnetic activity, drifts, electric field, or any other parameters for which a theoretical or empirical link to scintillation effects has been demonstrated.

At the present time the following bounds can be placed on the parameters:

Perturbation Strength  $C_s$  - Values of  $C_s$  as large as  $10^{22}$  in mks units are required to produce significant gigahertz scintillation. Work is currently underway to verify these values from in-situ data.

Spectral Index  $\nu$  - Scintillation data (particularly Wideband) show a  $\nu$  index with systematic variations, but values are generally less than 1.5 which corresponds to  $\kappa^{-2}$  one-dimensional in-situ spectral density function. In-situ data suggest a somewhat steeper slope than do the phase data. The discrepancy is not yet fully reconciled, although there are plausible reasons why it should exist.

Height and Thickness  $h, L$  - Scintillation is primarily an F-region phenomenon. Height certainly varies and must be carefully evaluated. Similarly, the effective thickness of scintillation producing regions can vary, particularly where narrow features are encountered.

Axial Ratios  $a, b$  - At the equator,  $a$  may be as large as 100. The evidence of a second preferred axis is uncertain at this time. In the auroral zone  $a \sim 10$  and in the nighttime precipitation regions  $b$  is finite and comparable to  $a$ , such that the irregularities are L-shell aligned sheets.

Cutoff Scales  $q_0, q_i$  - With regard to cutoff wavenumbers, the only safe statement is that the power-law continuum encompasses scale sizes from tens of kilometers to hundreds of meters. No systematic inner or outer scale cutoff has been reported from in-situ or scintillation data. However, local

input scales and diffusion cutoffs may well play an important role in the underlying physics.

### 3.4 Modeling of Phase and Intensity Scintillation, Including Geometrical Effects

Because of the increasing deployment of phase-sensitive radio systems that employ transionospheric propagation, the question has been raised whether a satisfactory phase scintillation model can be developed. In view of recent observations of phase-coherent satellite beacons, one may answer in the affirmative for two reasons.

The first reason is that the recent observations have attested to the utility of the phase-screen propagation model for calculating signal-statistical moments. While the details of a received phase pattern undoubtedly are altered by post-scattering propagation, the phase variance,  $\sigma_\phi^2$ , appears to change little. In particular,  $\sigma_\phi$  retains in expectation the proportionality to wavelength,  $\lambda$ , that one calculates in the near zone of a phase screen. (There does appear to be a slight decrease in the power law index of the phase spectrum with increasing Fresnel distance, however.) The recent observations also have demonstrated the viability of phase-screen calculations of the normalized variance,  $S_4^2$ , of signal intensity in terms of the ionosphericly imposed phase perturbations (Fremouw, 1979). Thus the calculations of  $\sigma_\phi$  performed in past modeling efforts and iteratively tested against (nonsaturated) observations of intensity scintillation should be as valid as the  $S_4$  values calculated from them.

The second contribution that the recent observations make to an affirmative answer is a high quality (albeit spatially and temporally limited) data base for direct testing of past and future models against measured values of  $\sigma_\phi$ . This will permit extending the models into regimes of saturated  $S_4$ .

The recent observations, however, have not supported the notion that simplistic extrapolation of  $S_4$  data would yield a viable phase-scintillation model. They have indicated, though, that phase-screen calculations are consistent with the observed behavior of the ratio  $S_4/\sigma_\phi$  so long as geometrical control of scintillation and the influence of both measurement procedures and operational systems are carefully accounted for.

A key factor in the production of scintillation is the large outer scale that the ionosphere imposes on the spatial spectrum of phase fluctuations. The upshot is that the phase spectrum is cut off at its low-frequency end by the detrending procedure employed in observations or by the duration of phase-sensitive performance (e.g., a coherent-integration period) in an operational system. In contrast, the spectrum of intensity scintillations is cut off by the Fresnel filtering that arises in the diffractive propagation process (which includes refractive focusing and defocusing as a special case).

For systems in which phase is effectively passed through a sharp-cutoff, high-pass filter, the phase variance of the scintillating signal is

$$\sigma_\phi^2 = \frac{2T}{p-1} f_c^{1-p} \quad (3)$$

where  $T$  = power-spectral density of phase at 1 Hz,

$p$  = power-law index of phase spectrum, and

$f_c$  = filter cutoff frequency (reciprocal of detrend period).



Rino (1979a,b) has developed an expression for  $T$  in terms of the power-spectral density,  $abC_s$ , of the three-dimensional spectrum of ionospheric electron density at an effective (nonisotropic) wavenumber of unity (1 rad/m), where  $a$  and  $b$  are axial ratios along and across the geomagnetic field. (For isotropic irregularities,  $a$  and  $b$  are unity, so the three-dimensional spectral density at 1 rad/m is  $C_s$  in that case).

The expression is

$$T = r_e^2 \lambda^2 (L \sec \theta) G N(\nu) v_e^{2\nu-1} C_s \quad (4)$$

where  $r_e$  = classical electron radius

$\lambda$  = radio wavelength

$\nu = p/2$

and  $N(\nu)$  is a normalizing factor that depends in a calculable way on the one-dimensional ionospheric power-law spectral index  $2\nu-1$  measured in situ. In Eq. (4),  $L$  is the thickness of the irregular layer and  $v_e$  is the effective velocity with which contours of equal correlation in the ionosphere are scanned.

The quantity  $G$  in Eq. (4) is a geometrical enhancement factor defined as

$$G = \frac{ab \sec \theta}{(AC-B^2/4)^{1/2}} \quad (5)$$

where  $A$ ,  $B$ , and  $C$  are functions of  $a$ ,  $b$ , and  $\delta$  (an angle prescribing the orientation of the sheetlike irregularities described by  $b \neq 1$ ) and of the incidence angle,  $\theta$ , and magnetic azimuth,  $\phi$ , of the radio propagation vector and the local magnetic dip,  $\psi$ .

For isotropic irregularities ( $a = b = 1$ ), the denominator in Eq. (5) reduces to  $\sec \theta$ , so  $G$  itself reduces to unity. For axially symmetric irregularities ( $a \neq b = 1$ ),  $G$  displays a fairly sharp peak when the propagation vector nears alignment with the elongation axis (i.e., the magnetic field direction). For sheet-like irregularities,  $G$  is enhanced whenever the line of sight grazes the sheets ( $L$  shells for  $\delta = 0$ ; magnetic meridian planes for  $\delta = 90^\circ$ ).

The effective velocity,  $v_e$ , contains the same factor in its denominator as that appearing in Eq. (5) because irregularities are scanned more rapidly when the line of sight is nearly aligned with an elongation axis. Thus more phase-fluctuation energy is contained within the high-pass spectral window employed in both observation and operation.

Combining Eqs. (3), (4), and (5), we find that the phase variance experienced by a system is

$$\sigma_\phi^2 = \left[ \frac{2r_e^2 \lambda^2 N(\nu)}{2\nu-1} \right] \{f_c^{(1-2\nu)}\} \{G \sec \theta v_e^{(2\nu-1)}\} \{LC_s\} \quad (6)$$

The first factor in brackets is essentially a constant for a given wavelength, although its estimation may be improved upon by measuring  $\nu$  either in-situ or from the phase spectrum. It is most important to note that the second factor in brackets depends upon the measuring system in modeling and upon the operational system in applications.



Besides estimation of  $\nu$ , modeling efforts should be directed at establishing the ionospheric structural parameters ( $a$ ,  $b$ , and  $\delta$ ) that dictate the geometrical behavior of scintillation as described by the third bracketed factor in Eq. (6), and the "height-integrated strength of turbulence,"  $LC_S$ .

Present best estimates for the parameters that describe the three-dimensional shape of ionospheric irregularities are

$$\nu \approx 1.5 (\approx 1.38 \text{ on average})$$

$$a \approx \begin{cases} 100 & \text{near the equator} \\ 10 & \text{elsewhere} \end{cases}$$

$$b \approx \begin{cases} 10 & \text{in the high-latitude particle precipitation zone} \\ 1 & \text{elsewhere} \end{cases}$$

$$\delta = 0 \text{ in the high-latitude particle precipitation zone}$$

There is some information on variations of  $\nu$ , but little is known about the behavior of the other parameters, which may be fairly simple, however. Two questions of interest are whether sheetlike irregularities exist near the equator and at middle latitudes and whether the precipitation-zone sheets extend into the polar caps from their known location in the region of diffuse (and probably discrete) auroras.

By far the most variable of the factors in Eq. (6) is  $LC_S$ . Most modeling effort should be put into modeling  $LC_S$  by means of scintillation data and in-situ measurements. It may also be possible to derive  $C_S$  from good quality ionograms in regions where frequency spread is correlated with scintillation. An unambiguous determination of  $LC_S$  from scintillation measurements requires compensation for geometrical effects while the in-situ technique requires an estimate of irregularity layer thickness  $L$ .

Rino (1979a,b) also has derived expressions for  $S_4$ . Indeed, for weak to moderate scintillation ( $S_4 \lesssim 0.8$ ), one may write the intensity variance per unit phase variance as

$$\frac{S_4^2}{\sigma_\phi^2} = N(\nu) \frac{F}{G} \left( \frac{f_c}{\nu_e} \right)^{2\nu-1} Z^{\nu-1/2} \quad (7)$$

where  $N(\nu)$  is a ratio of known normalization factors,  $F$  describes the geometrical behavior of the Fresnel filter, the next factor accounts for the fact that phase is spatially filtered by a temporal filter of cutoff  $f_0$  rather than by Fresnel filtering, and the final factor contains the usual Fresnel-zone size

$$Z = \frac{\lambda z \sec \theta}{4\pi} \quad (8)$$

Thus, once phase scintillation is modeled one may calculate the non-saturated intensity scintillation index by means of Eqs. (7) and (8), for which an estimate or model is needed for the height,  $h$ , of the irregularity layer. Under extreme conditions, one must resort to other approaches for calculating  $S_4$ . Alternatives are the two-component techniques of Fremouw and Rino (1978), which treat geometric-optics focusing directly and provides the full (joint) statistics of the complex signal, and Rino's (1979a, b) recently completed multiple-scatter theory.

For all scattering-strength regimes, what is needed is a geophysical/morphological model for the irregularity parameters (including their drift

velocity,  $V_d$ , for geostationary-satellite applications) as functions of the following:

$\ell_m$  = invariant latitude (dip latitude at the equator)

$\ell_o$  = longitude

$t$  = time of day

$D$  = day of the year

$R$  = a measure of epoch in the solar cycle

$X$  = one or more indicators of the level of ionospheric activity.

Currently existing morphological models could be significantly improved by means of recently acquired high-quality data bases, which include phase-scintillation measurements. For the improved models to provide predictive capability, it is the activity parameter  $X$  that must be identified and against which scintillation must be modeled.

The present models employ solar flux (Aarons et al., 1978) or sunspot number for  $R$ , primarily as a measure of epoch in the solar cycle. They also employ  $K_p$  as a kind of activity indicator for high-latitude scintillation. It is felt that AE and/or local magnetic  $K$  may provide a better measure of the departure of high-latitude scintillation severity from its expectation level (for a given set of the variables,  $\ell_m$ ,  $\ell_o$ ,  $t$ ,  $D$ , and  $R$ ). This contention is being investigated. For equatorial scintillation, a different parameter is required and may be identified by means of ongoing research.

Recent advances in relating AE and its component auroral electrojet indices, AL and AE, to physical characteristics of the solar wind and the interplanetary magnetic field provide encouragement in the quest for a predictive capability in high-latitude terms of scintillation models. Progress in understanding the production mechanism of scintillation-producing irregularities near the equator, the most descriptive of which appear to be triggered by the Rayleigh-Taylor instability, also bodes well for development of some predictive capability in the next few years.

### 3.5 Use of In Situ Irregularity Data in Phase and Amplitude Scintillation Modelling

The retarding potential analyzer (RPS) on board the orbiting satellites, OGO-6 and the Atmospheric Explorer series have provided ion-concentration (or electron concentration for charge neutrality) at F-region heights with an accuracy of 0.01% at a sampling rate as high as 200 per sec corresponding to a spatial resolution of 35 m (see Hanson et al., 1970 and 1973). The ion or electron concentration ( $N$ ) data is processed to obtain rms irregularity amplitude  $\Delta N/N$  over a few seconds time interval which corresponds to a few tens of km distance along the orbital track.

The electron density deviation  $\Delta N$  or alternatively the strength of turbulence parameter  $C_s$  (Rino, 1979a,b) derived from the above was utilized to develop a synoptic model of phase and amplitude scintillations in the equatorial region during the December and June solstices (Basu et al., 1976; Basu and Basu, 1979). This model is in agreement with ground-based scintillation measurements at widely spaced stations in the equatorial region (Aarons, 1977). The unlimited longitude coverage provided by in-situ measurements established a longitudinal control of equatorial scintillations rather than a seasonal dominance which was not accounted for in earlier scintillation models (Fremouw and Rino, 1973). It is to be recognized that modeling of scintillations from in-situ measurements requires an assumption



regarding the thickness of the irregularity layer, L. However, correlated radar and scintillation measurements in the equatorial region indicate that a layer thickness of 200 km is appropriate for modeling purposes (Basu et al., 1977, 1978).

The in-situ measurements thus provide a viable technique for phase and amplitude scintillation modeling as it directly provides a measure of  $C_S$  which sensitively controls the phase and amplitude scintillation indices. In this respect, it is important to note that in-situ measurements indicate a variation of  $C_S$  by as much as six orders of magnitude in the equatorial region.

With the high  $C_S$  values obtained from in-situ data, the occurrence of saturated UHF and moderate GHz scintillations near overhead positions at the magnetic equator can be modeled on the basis of existing theory (Basu and Basu, 1976). Scintillation modeling by in-situ measurements is not limited to the equatorial region and currently AE-C and AE-D data are being utilized to develop high latitude phase and amplitude scintillation models.

In addition to the morphological models discussed above, the existing high resolution (35 m) in-situ data are being processed to obtain power spectra of irregularity amplitude. This spectral information can be directly related to the power spectrum of phase scintillations as well as amplitude scintillations in the framework of diffraction theory for more sophisticated systems requiring signal structure information. It may also be possible to obtain information on the axial ratio of irregularities by performing in-situ measurements with satellites in equatorial and polar orbits such as AE-C and AE-E. It may be possible to provide such information on a near real-time basis if a suitable platform is available to make high resolution in-situ measurements.

### 3.6 Use of Spread-F Data in Scintillation Models

Field aligned irregularities in the F region are responsible for scintillations as well as some varieties of spread F as observed with ionosondes. In the middle and high latitudes these irregularities produce the frequency spreading component of spread F (Singleton, 1957). At these latitudes therefore, frequency spreading can be used to indicate values of  $\Delta N$ . It must be remembered, however, that  $\Delta N$  is proportional to  $\Delta f_o \cdot f_o$  and consequently, the background layer modulation via foF2 must be removed (Singleton, 1962, 1979).

At the equatorial latitudes, the field-aligned irregularities are substantially horizontal and hence give rise to a variety of range spreading, rather than classical mid-latitude frequency spreading. Further, this range spreading has characteristics different from those of mid-latitude range spreading (Rastogi, 1978). Also, at equatorial latitudes, it is primarily the range spreading which correlates with backscattering or scintillations while at middle latitudes this point needs further investigation (Singleton, 1963; Rastogi and Woodman, 1978a; Rastogi et al., 1976). All this complicates the situation and makes the spread F indicated on present ionospheric parameter tabulations of dubious usefulness for estimating  $\Delta N$  and, hence, for constructing scintillation models in the equatorial region. This difficulty, could be removed, at least as far as occurrence is concerned, by including in the tabulations (e.g., in association with one of the F-region height parameters) an indication that range spreading is occurring.



It is recommended that the known morphology of spread F (Singleton, 1960, 1968, 1975, 1977) be extended to take into account the different types of spread F, especially in the equatorial region. Wherever possible carefully calibrated ionograms should be analyzed to obtain the diurnal, seasonal, solar cycle and longitudinal variations of the different types of spread F.

There is little understanding of spread F irregularities in the topside ionosphere. The association of these irregularities with scintillations should be studied. Further, in the equatorial region, such studies should be continued to include the interrelation of plumes (Woodman and LaHoz, 1976) with scintillations and both bottomside and topside spread F.

The horizontal electric field in the F region appears to be the most promising candidate as a predictor of the occurrence of equatorial range spreading and scintillations. More investigations should be undertaken to isolate the effect this and other geophysical events have on spread F and scintillations (Rastogi and Woodman, 1978b).

#### 4. CONCLUSIONS AND RECOMMENDATIONS

##### 4.1 State-of-the-Art — Summary

In this working group report we have discussed the state of the art of both time delay and scintillation predictions. We have seen that, unless one is able to make a near real time measurement of TEC within a relatively small space-time cell, predictions are no better than the use of the monthly mean value. For amplitude scintillation, in large parts of the world, even the average climatology is unknown, and, for phase scintillation, preliminary models are emerging based on Wideband observations at a few locations and in-situ data with global coverage. System constraints, such as orbital path, data detrend time interval, etc., strongly dictate the magnitude of phase variance and, as such, any basic phase scintillation model has to be translated to suit a particular application.

Since the present state of transionospheric time delay and scintillation prediction is based largely upon empirical models there is a long-term need for a better understanding and application of theoretical considerations, which should lead to improved models. In the intermediate time frame, however, there is a need for greater spatial, as well as temporal, density of experimental data to improve the empirical models and to provide a firm data base to test new theoretical models.

As a result of our working group discussions we have made several recommendations and suggestions to improve existing techniques for transionospheric propagation predictions. These are detailed in the following section.

##### 4.2 Recommendations for New or Improved Data Sources

1. Appropriate radio beacon transmitters should be developed and placed on the new generation of operational Geostationary Meteorology Satellites to provide improved climatological information on both TEC and scintillation, as

well as to provide the opportunity for near real time measurements for improving predictions in specific areas.

2. The use of topside sounders for real time assessment of the ionosphere, both for topside TEC and topside spread F, should be investigated.

3. Signals from the NAVSTAR-Global Positioning Satellite should be considered for use for both TEC and scintillation real-time measurements. The proposed number and configuration of this satellite constellation offers potentially large geographic coverage for increased ionospheric measurement capability.

4. Studies of promising indirect methods of determining scintillation, such as in-situ monitoring of  $\Delta N$  at F-region heights, should be continued.

5. A program of routine, calibrated measurements of solar UV flux should be initiated and continued indefinitely for use in developing and improving empirical and physical models of the ionosphere.

#### 4.3 Suggestions for Additional Model Improvements

1. The behavior of TEC during magnetic storms should be studied in a more comprehensive set of longitude zones and also in the Southern Hemisphere.

2. Effort should be directed toward development of a more complete understanding of the temporal and spatial variability of measured effects; such as the scintillation S<sub>4</sub> index and TEC; that is, how does a perturbation from average conditions "track" in space and time?

3. The replacement of the present "source parameters" which drive the various transionospheric propagation models (i.e., K<sub>p</sub>, R<sub>z</sub>, or F<sub>10.7</sub>) by more meaningful parameters should be considered, based upon alternate parameter availability.

4. The effects of the variability of the neutral atmosphere on transionospheric propagation should be examined for inclusion in present models.

5. A comprehensive approach should be applied to transionospheric propagation problems, rather than the present piecemeal approach, such as concentrated sets of experiments that provide standardized data collection, calibration, assembling and archiving for use in model development and testing.

#### REFERENCES

[Those references marked with an asterisk (\*) are from the Solar-Terrestrial Predictions Workshop Proceedings, Boulder, Colorado.]

Aarons, J. (1977): Equatorial scintillations: A review. Trans. on Ant. and Prop., AP-25: 729.

Aarons, J., E. MacKenzie and K. Bhavnani (1978): Equatorial and high latitude empirical models of scintillation levels. In: Proc. of AGARD Conference #238 on Operational Modelling of the Aerospace Propagation Environment, Ottawa, Canada.

- Allen, R.S. (1977): Considerations relative to adapting TRANSIT Observations to predicting radar range corrections. AFGL-TR-77-0004, DDC# ADA 038238.
- Allen, R.S., D.E. Donatelli, G.K. Hartmann, R. Leitinger (1977): Adaptive mapping of mid-latitude Ionosphere, AFGL-TR-77-0176.
- Anderson, D.N. (1976): Modeling the mid-latitude F-region storm using east-west drift and a meridional wind. Planet. Space Sci., 24:69.
- Basu, S. and S. Basu (1976): Correlated measurements of scintillations and in-situ F-region irregularities from OGO-6. Geophys. Res. Lett., 3:681.
- Basu, S., S. Basu and B.H. Khan (1976): Model of equatorial scintillations from in-situ measurements. Radio Sci., 11:821.
- Basu, S., J. Aarons, J.P. McClure, C. LaHoz, A. Bushby and R.F. Woodman (1977): Preliminary comparisons of VHF radar maps of F-region irregularities with scintillations in the equatorial region. J. Atmos. Terr. Phys., 39:1251.
- Basu, S., S. Basu, J. Aarons, J.P. McClure and M.D. Cousins (1978): On the coexistence of kilometer- and meter-scale irregularities in the nighttime equatorial F region. J. Geophys. Res., 83:4219.
- \*Basu, S. and S. Basu (1979): Model of phase and amplitude scintillations from in-situ measurements.
- Bernhardt, P.A., D.A. Antoniadis, A.V. daRosa (1976): Lunar perturbations in columnar electron content and their interpretation in terms of dynamo electrostatic fields. J. Geophys. Res., 81:5957.
- Bernhardt, P.A. (1978): Identification of mixing frequency components in ionospheric electron content data by multiplicative homomorphic filtering. J. Geophys. Res., 83:5212.
- Bischoff, K. and B. Chytil (1969): A note on scintillation indices. Planet. Space Sci., 17:1059.
- Bramley, E.N. (1977): The accuracy of computing ionospheric radio-wave scintillation by the thin-phase-screen approximation. J. Atmos. Terr. Phys., 39: 39.
- Briggs, B.H. and I.A. Parkin (1963): On the variation of radio star and satellite scintillations with zenith angle. J. Atmos. Terr. Phys., 25:339.
- Costa, E. and M.C. Kelley (1976): Calculations of equatorial scintillations at VHF and gigahertz frequencies based on a new model of the disturbed equatorial ionosphere. Geophys. Res. Lett., 3:677.
- Davies, K., R.B. Fritz, R.N. Grubb and J.E. Jones (1975): Some early results from the ATS-6 radio beacon experiment. Radio Sci., 10:785.



- Davies, K., R.F. Fritz and T.B. Gray (1976): Measurements of the columnar electron contents of the ionosphere and plasmasphere. J. Geophys. Res., 81: 2825.
- Davies, K., D.N. Anderson, A.K. Paul, W. Degenhardt, G.K. Hartmann and R. Leitinger (1979): Nighttime increases in total electron content observed with the ATS-6 radio beacon. J. Geophys. Res., 84: 1536.
- Degenhardt, W., G.K. Hartmann and R. Leitinger (1977): Effects of a magnetic storm on the plasmaspheric electron content. J. Atmos. Terr. Phys., 39:1435.
- Donatelli, D.E. and R.S. Allen (1978): Temporal variability of ionospheric refraction correction. In: Effect of the Ionosphere on Space and Terrestrial Systems, Ed. J.M. Goodman, 490-496.
- \*Donatelli, D.E. and R.S. Allen (1979): Ionospheric refractive correction using an adaptive procedure.
- Elkins, T.J. and C.M. Rush (1973): A statistical predictive model of the polar ionosphere. AFCRL-TR-73-0331.
- Fleury, R. and J.P. Cornec (1978): Etude du contenu plasmaspherique et des scintillations a partir des signaux d'ATS-6 a Lannion. In: AGARD CP-238, Vol.1.
- Fremouw, E.J. and C.L. Rino (1973): An empirical model for average F-layer scintillation at VHF/UHF. Radio Sci., 8:213.
- Fremouw, E.J., R.L. Leadabrand, R.C. Livingston, M.D. Cousins, C.L. Rino, B.C. Fair and R.A. Long (1978): Early results from the DNA Wideband satellite experiment - complex signal scintillation. Radio Sci., 13:167.
- Fremouw, E.J. and C.L. Rino (1978): A signal-statistical and morphological model of ionospheric scintillation. In: Proc. of AGARD Conference #238 on Operational Modelling of the Aerospace Propagation Environment, Ottawa, Canada.
- \*Fremouw, E.J. (1979): Some results of the Wideband experiment pertinent to transionospheric communication-channel modeling.
- Fremouw, E.J., R.C. Livingston and D.A. Miller (1979): On the statistics of scintillation signals. Submitted to J. Atmos. Terr. Phys.
- Hajeb-Hosseini, H. (1977): Ionospheric studies using geostationary satellites. Thesis submitted to the University of Wales, Aberystwyth, Wales.
- Hanson, W.B., S. Sanatani, D. Zuccaro and T.W. Flowerday (1970): Plasma measurements with the retarding potential analyzer on OGO-6. J. Geophys. Res., 75:5483.

- Hanson, W.B., D.R. Zuccaro, C.R. Lippincott and S. Sanatani (1973): The retarding-potential analyzer on Atmosphere Explorer. Radio Sci., 8:333.
- Hartmann, G.K. (1978): The information explosion and its consequences for data acquisition, documentation and processing. Report UAG-65, World Data Center-A, U.S. Dept. of Commerce, Boulder, CO.
- Huang, Y.N. (1978): Lunar variations of total electron content at Lunping. J. Atmos. Terr. Phys., 40:1219.
- Jones, U.B., R.P. Graham and M. Leftin (1969): Advances in ionospheric mapping by numerical methods. ESSA Technical Reprint ERL 107-ITS-75.
- Jones, U.B. and D.L. Obitts (1970): Global representation and solar cycle variation of foF2 monthly median 1954-1958. Telecommunications Research Report. OT/ITS/TRR3.
- Kersley, L. and J.A. Klobuchar (1978): Comparison of protonospheric electron content measurements from the American and European sectors. Geophys. Res. Lett., 5:123.
- Kersley, L., H. Hajeb-Hosseini and K.J. Edwards (1978): Post-geomagnetic storm protonospheric replenishment. Nature, Lond., 271:430.
- Kersley, L. and J.A. Klobuchar: Storm associated protonospheric depletion and recovery. Unpublished manuscript.
- Klobuchar, J.A. and J.M. Johanson (1977): Correlation distance of mean daytime electron content. AFGL-TR-77-0185, DCA# ADA 048117.
- Klobuchar, J.A., M.J. Buonsanto, M.J. Mendillo and J.M. Johanson (1978): The contribution of the plasmasphere to total time delay. In: Effect of the Ionosphere on Space and Terrestrial Systems, Ed. J.M. Goodman, 486-489.
- \*Lassudrie-Duchesne, P., A.M. Bourdila, and H. Sizun (1979): The French short term radiopropagation predictions in the decameter band.
- Leitinger, R., R.S. Allen, D.E. Donatelli and G.K. Hartmann (1978): Adaptive mapping of ionospheric features. In: Effect of the Ionosphere on Space and Terrestrial Systems, Ed. J.M. Goodman, 530-537.
- \*McNamara, L.F. (1979): The use of ionospheric indices to make real-and near-real-time forecasts of foF2 around Australia.
- Mendillo, M. (1973): A study of the relationship between geomagnetic storms and ionospheric disturbances at mid-latitudes. Planet. and Space Sci., 21:349-358.
- \*Mendillo, M., and J.A. Klobuchar (1979): A morphology-based prediction scheme for the coupled latitudinal and local-time development of F-region storms.

- Mulkern, F.A. (1976): Comparison of predicted monthly medians of total electron content with field observations. AFGL-TR-76-0158.
- Poletti-Liuzzi, D.A., K.C. Yeh, and C.H. Liu (1977): Radio beacon studies of the plasmasphere. J. Geophys. Res., 82:1106-1114.
- Prolss, G.W. (1977): Seasonal variations of atmospheric-ionospheric disturbances. J. Geophys. Res., 82:1635-1640.
- Prolss, G.W., and U. von Zahn (1977): Seasonal variation in the latitudinal structure of atmospheric disturbances. J. Geophys. Res., 82(35):5629-5632.
- Rao, N.N., and C.F. Stubenrauch (1967): Lunar tidal variations in the equivalent slab thickness of the ionosphere over Hawaii. J. Geophys. Res., 72:5547.
- Rastogi, R.G., M.R. Deshpande, A. Sen, K. Davies, R.N. Grubb, and J.E. Jones (1976): Proc. Symp. Geophysical Use of Satellite Beacon Observations, Boston University, Boston, Mass., p. 518.
- Rastogi, R.G., and R.F. Woodman (1978): Spread-F in equatorial ionograms associated with reversal of horizontal F-region electric field. Ann. Geophys., 34:31-36.
- Rastogi, R.G., and R.F. Woodman (1978): VHF radio wave scattering due to range and frequency types of equatorial spread-F. J. Atmos. Terr. Phys., 40(4):485.
- \*Reddy, B.M., S. Aggarwal, D.R. Lakshmi, S. Sastry, and A.P. Mitra (1979): Long term solar activity and ionospheric prediction services rendered by National Physical Laboratory.
- Rino, C.L., and E.J. Fremouw (1977): The angle dependence of singly scattered wavefields. J. Atmos. Terr. Phys., 39:859.
- Rino, C.L. (1979a): A power-law phase screen model for ionospheric scintillation. (1) Weak scatter. Accepted for publication, Radio Sci.
- Rino, C.L. (1979b): A power-law phase screen model for ionospheric scintillation. (2) Strong scatter. Accepted for publication, Radio Sci.
- Rishbeth, H., and H. Kohl (1976): Topical questions of ionospheric physics: a working group report. J. Atmos. Terr. Phys., 38:775-780.
- Rufenach, C.L. (1975): Ionospheric scintillation by a random phase screen: spectral approach. Radio Sci., 10:155.
- Rush, C.M. (1976): An ionospheric observation network for use in short-term propagation predictions. Telecommunication Journal, 43(VIII):544-549.



- Ryan, D. (1979): Evaluation of the bias of a model of monthly mean TEC. Manuscript in preparation.
- Singleton, D.G. (1957): A study of spread-F ionospheric echoes at night at Brisbane, 3, frequency spreading. Aust. J. Phys., 10:60.
- Singleton, D.G. (1960): The geomorphology of spread F. J. Geophys. Res., 65(11):3615.
- Singleton, D.G. (1962): Spread-F and the perturbations of the maximum electron density of the F layer. Aust. J. Phys., 15:242-260.
- Singleton, D.G. (1963): The spread-F equator. J. Atmos. Terr. Phys., 25(3): 121.
- Singleton, D.G. (1968): The morphology of spread-F occurrence over half a sunspot cycle. J. Geophys. Res., 73(1):295.
- Singleton, D.G. (1975): An empirical model of global spread-F occurrence. J. Atmos. Terr. Phys., 37(12):1535.
- Singleton, D.G. (1977): The reconciliation of an F-region irregularity model with sunspot cycle variations in spread-F occurrence. Radio Sci., 12(1): 107.
- \* Singleton, D.G. (1979): An improved ionospheric irregularity model.
- Soicher, H. (1976): Response of electrons in ionosphere and plasmasphere to magnetic storms. Nature, London, 259:33-35.
- Soicher, H. (1976): Diurnal, day-to-day, and seasonal variability of  $N_F$ ,  $N^+$  and  $N_p$  at Fort Monmouth, New Jersey. In: Cospar Satellite Beacon Symposium Proceedings, edited by M. Mendillo, Boston University, Boston, Mass., June.
- Webb, D.C., and L.J. Lanzerotti (1977): Temporal variations in total equatorial plasmaspheric content. J. Geophys. Res., 82:5201.
- Whitney, H.E., C. Malik, and J. Aarons (1969): A proposed index for measuring ionospheric scintillation. Planet. Space Sci., 17:1069.
- Whitney, H.E. (1974): Notes on the relationship of scintillation index to probability distributions and their uses for system design, AFCRL-TR-74-0004, Hanscom AFB, MA, ADA 778092.
- Whitney, H.E., and S. Basu (1977): The effect of ionospheric scintillation on VHF/UHF satellite communications. Radio Sci., 12:123-133.
- \*Wilkinson, P.J. (1979): Prediction limits for foF2.
- Woodman, R.F., and C. LaHoz (1976): Radar observations of F-region equatorial irregularities. J. Geophys. Res., 81:5447.

Yeh, K.C., C.H. Liu, and M.Y. Youakim (1975): A theoretical study of the ionospheric scintillation behavior caused by multiple scattering. Radio Sci., 10:97-106.

\*Zevakina, R.A., and E.V. Lavrova (1979): On the possibility to predict variations in the F2 region parameters as a function of the IMF direction.

Participants in Working Group on Communications Predictions  
Subsection on Transionospheric Propagation Predictions

Santimay Basu  
Physics Research Division  
Emmanuel College  
400 The Fenway  
Boston, MA 02115

J.A. Klobuchar (Subsection Leader)  
Air Force Geophysics Laboratory  
Hanscom AFB  
Bedford, MA 01731

Sunanda Basu  
Physics Research Division  
Emmanuel College  
400 The Fenway  
Boston, MA 02115

R. Leitinger  
Institute for Meteorology and  
Geophysics  
University of Graz  
Graz, A-8010, Austria

P.A. Bernhardt  
Radio Science Laboratory  
Stanford University  
Stanford, CA

M.J. Mendillo  
Astronomy Department  
Boston University  
Boston, MA 02215

K. Davies  
NOAA/Space Environment Lab.  
Boulder, CO 80302

M. Pokempner  
1-3413  
Institute for Telecommunications  
Sciences  
Boulder, CO 80303

D.E. Donatelli  
Research Center  
Regis College  
Weston, MA

R.G. Rastogi  
Air Force Geophysics Laboratory  
Hanscom AFB, Bedford, MA 01731

E.J. Fremouw  
Physical Dynamics, Inc.  
P.O. Box 3027  
Bellevue, WA 98009

C.L. Rino  
Stanford Research Institute  
Menlo Park, CA 94035

J.M. Goodman  
Space Environment Branch  
Code 7950  
Naval Research Lab.  
Washington, D.C. 20375

J. Secan  
Air Force Global Weather Central/TSIS  
Offutt AFB, NE 68113

G.K. Hartmann  
Max-Planck Institute for Aeronomy  
3411 Katlenburg  
Lindau 3,  
Federal Republic of Germany

D.G. Singleton  
Weapons Research Establishment  
GPO Box 2151  
Adelaide, S.A. 5001  
Australia

H. Soicher  
US Army Communications R&D  
Attn: DRDCO-COM-RF-5  
Fort Monmouth, NJ 07703

Note: The following were not able to attend the working group meeting, but made major contributions to the success of the working group efforts by correspondence.

J. Aarons  
Air Force Geophysics Laboratory  
Hanscom AFB  
Bedford, MA 01731

K.C. Yeh  
Department of Electrical Engineering  
University of Illinois  
Urbana, Illinois 61801



## IV. SOLAR ACTIVITY PREDICTIONS

### A. LONG-TERM SOLAR ACTIVITY

A WORKING GROUP REPORT prepared by P. S. McIntosh, Chairman, Members:  
G. M. Brown, R. Buhmann, T. Clark, P. F. Fougere, H. Hunter, J. V. Lincoln,  
H. H. Sargent III, J. G. Timothy, and Y.-Z. Lin

#### 1. TEMPORAL AND SPATIAL STRUCTURE OF SOLAR ACTIVITY

##### 1.1 Cautionary Remarks

This report of the working group on long-range solar activity predictions is not a comprehensive summary of the field. The membership of the working group did not represent all facets of this field; therefore, the chairman has taken liberty to discuss some ideas and results that did not arise during the conference. This was done as partial compensation for the lack of a general review paper on long-range predictions. The reader should beware that much more could be added to what will follow, and it is expected that some will object to the apparent ranking of results and to other implied value judgements.

##### 1.2 Need for Long-Term Predictions

The need for solar activity predictions beyond the next day or two is stated emphatically by nearly every working group in this conference. The need appears to be growing as our society builds more advanced technologies, makes more extensive use of the space environment and tries to plan for variations in climatic conditions.

Some of the stated needs for long-term predictions may not mention solar activity directly, but may state a requirement to know interplanetary conditions or conditions in the Earth's upper atmosphere. In nearly every case, predictions of these conditions are predicated on a prediction of solar activity, or prediction of large-scale solar structures that control or create the interplanetary conditions that affect the Earth.

The need for long-range solar predictions can be expressed in terms of the economics of communications and space systems design and the costs of inefficient or ineffective exploration for natural resources. Research in the fields of solar physics, space physics and geophysics can also become more efficient and effective through successful long-range solar activity predictions.

Radiation effects on both hardware and astronauts is of increasing concern. Long-range predictions of solar radiations can reduce the costs and risks of protecting against radiation effects.

The effects of climatic change on food production and fuel use are forcing attention to long-range prediction of climate. Evidence is accumulating that long-term solar variability is an important factor affecting climate.

### 1.3 Long-Term Time Scales

The prediction of solar activity more than three days in advance is the basic definition of the area of long-term solar activity predictions. This is simply any time frame longer than the present standard predictions issued by the U. S. Space Environment Services Center and other comparable solar-terrestrial services. Long-term solar predictions can be divided into six time scales: (1) the period from three days to two weeks that includes the evolution and activity of features present on the visible solar hemisphere; (2) 27 days in advance and multiples of this, making use of the long-term persistence of active centers and large-scale coronal structures and their periodic alignment with the Earth with the period of the solar rotation; (3) quasi-periodic pulses of activity that are superposed on the mean curve of the solar cycle, with periods varying from 3 to 5 solar rotations for secondary fluctuations and from 8 to 17 solar rotations for primary fluctuations (Dodson and Hedeman, 1970); (4) the 11-year cycle of solar activity variations that has been characteristic of the last two centuries; (5) the 22-year cycle in magnetic polarities and in the source of recurrent geomagnetic storms; and (6) very-long term variations in the level of solar activity, involving centuries and marked by minima such as that recorded as the Maunder Minimum near the end of the 17th Century.

### 1.4 Spatial Organization of Solar Activity

#### 1.4.1 Observations

Observations of temporal periodicities in solar activity certainly encourage attempts to predict that activity, but such predictions often take on a purely empirical aspect with no consideration of possible physical causes for the variations. Evidence for a physical basis for solar predictions comes from the observed spatial regularities in solar phenomena.

Bumba and Howard (1969) were among the first to recognize that the large-scale solar magnetic fields occurred in patterns that persist for many months, forming "rows" and "streams" in their diagrams of a time series of solar synoptic charts. These giant, regular structures are clearly seen in the solar charts constructed from H-alpha observations (McIntosh, 1972, 1975, 1979) and show even more persistence than in the magnetographic observations. Large-scale equatorial features can be identified for as long as two years, and features at high solar latitudes have been identified for more than four years. It follows that such persistence implies a measure of long-term predictability to large-scale aspects of solar activity.

The solar phenomena that appear to be organized by the large-scale solar magnetism are solar wind, solar particles and the strong activity complexes that produce large flares and proton events. The three-dimensional coronal



structures observed in X-rays conform closely to the "neutral" lines that form the boundaries to the large-scale solar magnetic patterns (McIntosh et al., 1976). The H-alpha maps of neutral lines provide the clearest and most distinctive pattern corresponding to the location of long-lived coronal holes (McIntosh, 1976, 1979). Recurrent geomagnetic storms are now linked to long-lived, high-speed solar wind streams that originate over coronal holes (Zirker, 1977).

It is well known that solar particles do not arrive at Earth from all large flares but show a strong preference for flares occurring in the western half of the solar disk. This simply implies that the solar magnetic fields, stretched into interplanetary space by the solar wind, are controlling the propagation of solar particles. The detailed time dependence of solar particle flux has been tediously compared with the large-scale magnetic structures surrounding the flaring sites, with correspondence in some detail between the two data (Roelof and Krimigis, 1973; Roelof, 1974).

Solar activity displays a strong longitude dependence when only the greatest sunspots and flares are included in the statistics (Bumba and Obridko, 1969). The longitude dependence of proton-producing flares is so striking that Svestka (1968) was led to suggest that solar predictions should make use of this observation. Calcium plage shows organization into long-lived row systems similar to, or even identical to, the patterns found for solar magnetic fields, and activity appears enhanced in regions where adjacent rows cross (Ambroz, 1973). Convergence and divergence among the long-lived magnetic-field patterns has also been noted by McIntosh (1972, 1976, 1979), and so it may be that the approach and merger of long-lived structures are associated with the formation of activity complexes that persist for more than one solar rotation.

#### 1.4.2 Implications for Predictions

The presence of stable patterns of solar activity and large-scale coronal structures immediately provides some assistance in predicting future solar activity. It is this stable, longitude dependent aspect of solar activity that makes for 27-day recurrence in several of the solar indices. Beyond simply repeating the observed 27-day pattern as a forecast for the next 27 days, the drifts of patterns over several solar rotations could be extrapolated to slightly alter this 27-day forecast.

Studies of the dynamics of large-scale magnetic patterns have been made exclusively with Mt. Wilson magnetic field data and only for two solar cycles during a period when the data quality has been evolving. The evolution of the data itself and the limitation to just two solar cycles makes it difficult to obtain definitive results. The fact that only a few workers have devoted themselves to the question of large-scale magnetism also imposes additional limitations. We are hopeful that the independent view of large-scale magnetism coming from the new atlas of H-alpha synoptic charts can contribute significant new information. Specifically, the improved resolution to the boundaries of the large-scale patterns may provide an ability to measure motions more precisely than previously.



## 1.5 Ancient Evidence for Solar Variability

The evidence for the very long-term solar variations comes from a number of sources that pre-date the modern era of scientific investigation and record keeping, and so is somewhat uncertain, precluding any serious consideration of periods in this long-term record. The Chinese delegates brought records of naked-eye sunspots from 43 B.C. and pointed out the occurrence of no fewer than seven naked-eye sunspots during the time of the Maunder Minimum. S. McKenna-Lawlor has examined Irish annals dated as early as 400 A.D. that include records of natural phenomena that may be interpreted as auroral occurrences. Study of these records may improve our understanding of long-term solar variations.

The fact that there is evidence for very-long term solar variations warns us not to put high confidence in predictions of solar cycles more than one cycle in advance. Predictions for at least one 11-year solar cycle ahead are encouraged, however, by the high degree of regularity with which a cycle of near this period has characterized the sunspot record of the past 250 years.

## 1.6 Definition of Solar Minimum

The view of the solar activity cycle as displayed in a plot of sunspot numbers has led to the prevailing impression that a cycle begins at the time of sunspot number minimum. We have known for some time that the new solar cycle actually overlaps the old by several years. Sunspot groups of the new cycle appear at mid solar latitudes at least a year before the minimum in sunspot number, while sunspots of the old, declining cycle continue near the solar equator as long as a year after sunspot number minimum. More recently, however, it has become apparent that the earliest manifestations of a new solar cycle occur significantly earlier than the sunspot minimum epoch. For example, the polar magnetic-field polarity reverses approximately two years after the peak of the preceding solar cycle. Dynamo theories of the solar cycle postulate that these polar fields represent the poloidal magnetic field system that is converted to a toroidal field by the dynamo process, amplifying the field to create the active-region magnetic fields of the new solar cycle. From the prediction point of view, such precursors to a new cycle, occurring prior to its conventional beginning at sunspot minimum, provide important physical backing to support a variety of recent prediction techniques based on the concept that the magnitude of the forthcoming cycle is predetermined at or before the time of sunspot minimum.

Following the polar magnetic-field reversal, the entire sun appears to undergo a reorganization of its large-scale structures as well as a change in the nature of its differential rotation. Large unipolar features develop which are open in the overlying corona and form coronal holes. Neither the coronal holes nor the underlying magnetic patterns undergo the deformation with time that would be predicted by the standard formulae of differential rotation. Out of these stable, long-lived coronal holes comes a large fraction of the high-speed solar wind streams that are the primary cause of the recurrent geomagnetic disturbances. The use of the index of recurrent

geomagnetic activity as a predictor of the magnitude of the next solar maximum (see following Section) implies that the pre-minimum solar global stability in some way reflects the mechanism that will build the next solar cycle.

Further, preliminary evidence for the solar cycle having its beginning earlier than the present definition of the solar minimum is suggested in the recent satellite measurements of solar EUV flux. A minimum in the EUV flux is reported to have occurred in 1975, whereas the minimum in the smoothed sunspot number occurred in mid-1976.

## 2. SOLAR CYCLE PREDICTION METHODS

Methods of predicting sunspot numbers fall into five distinct categories discussed below. The inherent accuracy of the methods varies considerably, but a typical error bar would be  $\pm 20\%$ .

### 2.1 Statistical Treatment of Past Solar Cycles

These methods depend on the existence of periodicities in the time series of sunspot numbers and use a variety of statistical approaches and spectral analysis techniques. In principle some of these methods can predict the progress of the entire next solar cycle, but with larger-than-average error bars. Some methods are amenable to projections further into the future, building on the prediction of the forthcoming cycle, but the uncertainties clearly accumulate rapidly. Several methods appear to predict up to one year into the future with higher-than-average accuracy.

### 2.2 Phenomena During the Previous Cycle

These approaches use solar and geophysical data appropriate to one cycle, and particularly during its declining phase, to predict the magnitude of the peak of the next cycle. Most of these methods may be applied around the time of sunspot minimum so that a lead time of at least two to three years is possible in the prediction of the next maximum. Some methods appear to have a higher-than-average potential accuracy. Consider the following examples of data used: (a) Magnetic activity, mainly recurrent, during the declining cycle (individual methods use the activity variously over the whole declining phase, the activity over the three years preceding solar minimum, or the magnitude of the peak (recurrent) activity during the declining phase). (b) Recurrent auroral activity over the declining phase. (c) Occurrence of "abnormal quiet days" (showing anomalous diurnal variations in the geomagnetic field) during the declining cycle. (d) Skewness of the time-rate-of-change of sunspot number for the preceding cycle. See papers by Ohl and Ohl and by Brown following this report.



### 2.3 Phenomena Around Sunspot Minimum

Closely allied to the methods of the preceding category are methods employing phenomena whose behavior near the time of solar minimum appears to relate to the magnitude of the subsequent sunspot maximum. These methods offer a similar lead time of at least two to three years. Examples of data used are (a) magnitude of minimum magnetic activity, and (b) magnitude of minimum sunspot number, (c) latitude of the first new cycle sunspots, (d) magnitude of the solar polar magnetic field, or (e) occurrence of long-lived away-sectors in the interplanetary magnetic field.

### 2.4 Initial Progress of the New Cycle

Two methods are based on a direct connection between the rate of increase of the new cycle and the magnitude of the peak subsequently attained. For these it is necessary to use data over the first 1 1/2 to 2 years into the cycle in question, giving a lead time of about two years in predicting the size of the next sunspot maximum. Predictions by the first method can be progressively updated as new data come in; this is not the case with the second method which appears to have a higher-than-average potential accuracy. Examples of data used are (a) the rate of rise of sunspot number, or (b) the rise of sunspot areas relative to faculae areas.

### 2.5 Influence of Planetary Tidal Action

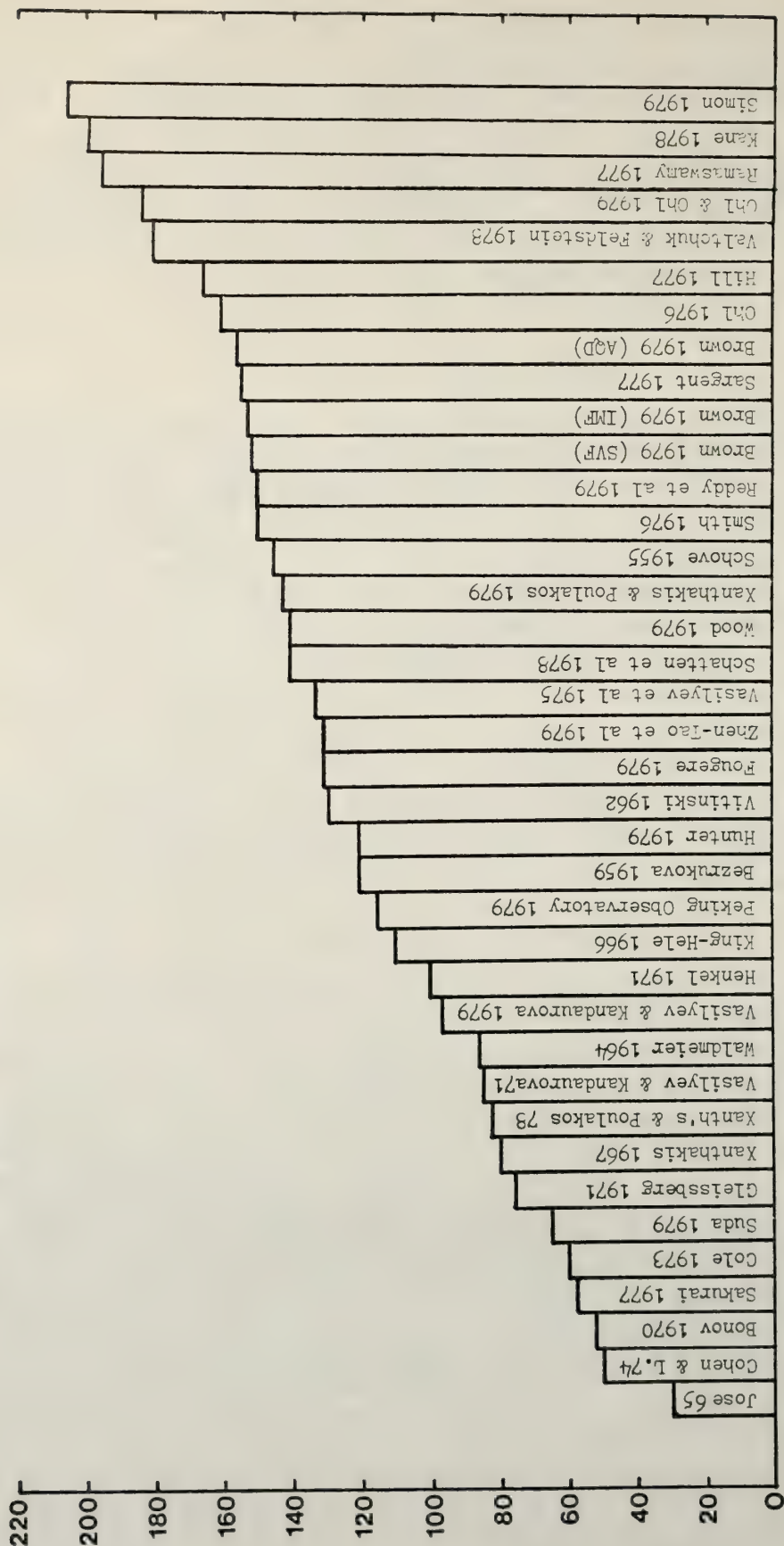
Analyses of the periodicities contained in the record of sunspot numbers continue to reveal periods remarkably close to those between various alignments of the planets. These relationships, however, have yet to be convincingly established before they can be used to make reliable predictions of solar activity.

## 3. Accuracy of Sunspot Cycle Predictions

It is possible to evaluate a confidence interval for a given prediction method, based on the individual approach used. This may translate into an inherent prediction accuracy of typically 10 or 20 percent. However, it is not possible with the present limited data to extrapolate from inherent accuracies to estimates of the real prediction accuracy of sunspot number. It is felt that this will only be possible after comparison of observed values with predictions for at least 500 years of reliable data, and probably for 1500-2000 years! There are so many predictions available for the coming maximum of Solar Cycle 21 (see Fig. 1) that several with similar values have been derived by different methods. Confirmation of particular predictions by the observed sunspot values will not therefore discriminate among several different prediction schemes. Only after each method has been used for a number of cycles can some be eliminated; hence, for the near future other criteria are needed for selecting realistic prediction methods.



# CYCLE 21 PREDICTED MAXIMUM



#### 4. Long-Term Predictions of Great Solar Events

The most severe terrestrial disturbances and the largest increase in the interplanetary radiation hazard follow the occurrence of the very rare energetic flares. There are presently no techniques available to anticipate these events prior to the appearance of their associated sunspots. This working group recommends more extensive research into the secular variation of the character of active regions from cycle to cycle, seeking to understand why the frequencies of naked-eye spot groups, great geomagnetic storms and ground-level cosmic-ray events do not correlate with the magnitude of the solar cycle as measured by sunspot number. We infer that there may be important changes in the distribution of strong solar magnetic fields that are not reflected in current solar activity indices. Inasmuch as great sunspot groups often occur in complexes of active regions which, in turn, occur in preferred longitudes, we suspect that fundamental changes in global solar magnetic fields are related to the occurrence of the greatest flares.

#### 5. Summary and Future Prospects

Our present capability to predict solar activity beyond the present solar rotation period (27 days) is limited to the following: (1) qualitative prediction of the solar locations where activity will be above (or below) the average for the current phase of the solar cycle, for a period of one or two solar rotations into the future; (2) qualitative prediction of whether activity in the next two months will subside or increase over the average level of the past two months; (3) quantitative predictions of the magnitude and time of occurrence of the next solar cycle maximum to within 20% of the observed value, with lead time of up to five years before its time of occurrence.

It is generally accepted that predictions beyond this time frame are unreliable for use in solar-terrestrial applications. Prediction of the time of solar maximum is a by-product of the prediction of the magnitude of maximum, using one of the accepted inverse relations between rise-time and peak magnitude.

It is likely that further progress in long-term prediction techniques will come about during the next ten years through improvements of our understanding of the processes involved in the development of solar magnetic fields throughout the solar cycle, and through study of the coupling of solar fields and particles into the terrestrial magnetosphere and ionosphere. Techniques are now being introduced to monitor the large-scale magnetic field structures in the photosphere and to model the extension of these magnetic fields into the corona. Satellite observations at XUV wavelengths can be employed directly to monitor the large-scale coronal magnetic field structures through their manifestation as coronal holes. The strong relationship between geomagnetic indices during the declining phase of the solar cycle and the magnitude of the succeeding cycle indicates the importance of investigating relationships between large-scale magnetic-field structures and the development of the localized active-region magnetic fields. Along these lines it seems possible that a coronal-hole index may replace the currently-used geomagnetic indices to provide a more accurate prediction of the magnitude of the succeeding solar cycle.

It is also possible that the absolute magnitude of the solar EUV flux may provide additional data on the state of development of an activity cycle. If, as has been reported, the EUV minimum precedes the sunspot number minimum, the EUV data may be used to predict the date of the sunspot number minimum and, thereby, increase the accuracy of the prediction data for the coming activity cycle.

Since the source of the EUV radiation, and of the particle fluxes produced during flares, lies in the chromosphere and transition-region, direct imaging of these layers of the solar atmosphere may well allow us to determine those characteristics of centers of activity that lead to major flare activity and determine the resulting level of geomagnetic activity.

Some recent developments in solar physics which may be relevant to improvements in long-range solar predictions include: (1) regular measurement of large-scale solar velocity fields; (2) discovery that the rate of equatorial solar rotation was faster prior to the Maunder Minimum than currently; (3) discovery that the solar differential rotation equation is variable, possibly as a function of phase of the solar cycle, and differing between the two solar hemispheres; (4) discovery of large-scale magnetic-field patterns with lifetimes as much as half that of the 11-year solar cycle, some of which correlate with the positions of coronal holes; (5) discovery of large-scale convergence and divergence among magnetic patterns, with a probable role in the formation of coronal holes; and (6) discovery of global solar oscillations.

## 6. Recommendations

The Working Group on Long-Term Solar Activity Predictions recommends the following:

1. Synoptic solar observations should be renewed, modernized and researched because of their fundamental importance to a full understanding of the solar processes leading to active-region formation and evolution and to the control of solar emanations that alter the terrestrial environment. We emphasize the need to know relationships between slowly-varying large-scale aspects of solar activity and the short-term, localized sources of important solar flares.

2. The existing long-term records of solar activity should be continued without interruption in order to maintain a self consistent historical record of variations in the level of solar activity. It is of paramount importance to continue at the very minimum: (a) the Zurich sunspot number, which is the only standardized index of solar activity covering nearly two centuries in time, and (b) the Ottawa 10.7 cm radio-noise index.

3. Photoheliographic recordings of sunspot positions, areas and classifications and measurements of solar faculae are critically needed and should be continued. Such recordings have recently been terminated by the Greenwich Observatory. It is crucial that such records continue with the same high standards of calibration and data reduction.



4. The indices and measurements of geomagnetic activity, which have a demonstrated relationship to the level of solar activity in the succeeding solar cycle, should be maintained as prediction tools. These indices or measurements include, but are not necessarily restricted to (a) aa, (b) AE, (c) Kp and (d) the H component of the magnetic field.

5. New indices of solar activity should be developed that utilize measurements from space. There is a critical need for (a) measurements of the solar EUV and X-ray fluxes for a direct determination of the variation of the energy input into the terrestrial atmosphere; and possibly also a prior indication of the occurrence of solar minimum; and (b) images of the solar disk at EUV and X-ray wavelengths in order to monitor the presence of large-scale magnetic field structures which can be utilized during the declining phase of a cycle as predictors of the level of activity during the succeeding cycle.

## 7. Acknowledgements

This report is substantially the opinions of the chairman and he takes responsibility for all its imperfections. Special thanks are deserved by G. M. Brown, J. G. Timothy, T. Clark, J. Virginia Lincoln, and H. H. Sargent III for their efforts toward completing this report, both during and after the conference.

## 8. References

- Ambroz, P. (1973): Structural changes and regularities in the distribution of calcium flocculae on the solar surface during solar cycle 19. Bull. Astron. Inst. Czech., 24:80.
- Bezrukova, A. Y. (1959): Soln. Dann., No. 11.
- Bonov, A. D. (1970): Soln. Dann., No. 7, 111.
- Brown, G. M. (1979): Solar-Terrestrial Predictions Proceedings, Vol. II.
- Bumba, V. and Howard, R. (1969): Solar Phys., 7:28.
- Bumba, V. and Obridko, V. N. (1969): Bartels' "active longitudes", sector boundaries and flare activity. Solar Phys., 6:104.
- Cohen, T. J. and Lintz, P. R. (1974): Nature, 250, 398.
- Cole, T. W. (1973): Solar Phys., 30:103.
- Fougere, P. F. (1979): Solar-Terrestrial Predictions Proceedings, Vol. III.
- Gleissberg, W. (1971): Solar Phys., 21:240.
- Henkel, H. R. (1971): Solar Phys., 20:345.

- Hill, J. R. (1977): Nature, 266:151.
- Hunter, H. E. (1979): Solar-Terrestrial Predictions Proceedings, Vol. III.
- Jose, P. D. (1965): Astron. J., 70:193, and Henkel, 1971 (q.v.).
- Kane, R. P. (1978): Nature, 274:139.
- King-Hele, D. G. (1966): Nature, 209:285.
- McIntosh, P. S. (1972): Inference of solar magnetic polarities from H-alpha observations. Solar Activity Observations and Predictions, P. S. McIntosh and M. Dryer, eds., Vol. 30 of Progress in Astronautics and Aeronautics series, Martin Summerfield, series editor, The MIT Press, 65-92.
- McIntosh, P. S. (1975): H-alpha synoptic charts of solar activity for the period of Skylab observations. Report UAG-40, World Data Center A for Solar-Terrestrial Physics, Boulder, Colo.
- McIntosh, P. S. (1976): An empirical model for the origin and evolution of coronal holes. Amer. Astron. Soc. Bull., 8:325.
- McIntosh, P. S., Krieger, A. S., Nolte, J. T., and Vaiana, G. (1976): Association of X-ray arches with chromospheric neutral lines. Solar Phys., 49:57.
- McIntosh, P. S. (1979): Large-scale solar magnetic fields during solar cycle 20. Proceedings of IAU Symposium No. 91, D. Reidel, in press. M. Dryer and E. Tandberg-Hanssen, Editors.
- Ohl, A. I. (1976): Soln. Dann., No. 9, 73.
- Ohl, A. I. and Ohl, G. I. (1979): Solar-Terrestrial Predictions Proceedings, Vol. II
- Peking Observatory (1979): Solar-Terrestrial Predictions Proceedings, Vol. I
- Ramaswamy, G. (1977): Nature, 265:713.
- Reddy, B. M., Aggarwal, S., Lakshmi, D. R., Sastry, S., and Mitra, A. P. (1979): Solar-Terrestrial Predictions Proceedings, Vol. I.
- Sakurai, K. (1977): Nature, 269:401.
- Sargent, H. H. (1977): Trans. Amer. Geophys. Union, 58:1220.
- Schatten, K. H., Scherrer, P. H., Svalgaard, L., and Wilcox, J. M. (1978): Geophys. Res. Letters, 5:411.
- Schove, D. J. (1955): J. Geophys. Res., 60:127.
- Simon, P. A. (1979): Solar Phys. (in press).

Smith, F. M. (1976): Radio Comm., 52:494.

Suda, T. (1979): Solar-Terrestrial Predictions Proceedings, Vol. III.

Svestka, Z. (1968): On long-term forecasts of proton flares. Solar Phys., 4:18.

Valtchuk, T. E. and Feldstein, Y. I. (1978): Soln. Dann., No. 12.

Vasilyev, O. B. and Kandaurova, K. A. (1971): Soln. Dann., No. 11, 109.

Vasilyev, O. B. and Kandaurova, K. A. (1971): Solar-Terrestrial Predictions Proceedings, Vol. III.

Vasilyev, O. B., Vitinski, Y. I., and Kandaurova, K. A. (1975): Soln. Dann., No. 10, 55.

Vitinski, Y. I. (1962): Solar-Activity Forecasting, Acad. Nauk. SSSR, Leningrad.

Waldmeier, M. (1964): In Xanthakis and Poulakos, 1978 (q.v.).

Wood, C. A. (1979): Solar-Terrestrial Predictions Proceedings, Vol. III.

Xanthakis, J. N. (1967): Nature, 215:1046.

Xanthakis, J. N. and Poulakos, C. (1978): Solar Phys., 56:467.

Xanthakis, J. N. and Poulakos, C. (1978): Solar-Terrestrial Predictions Proceedings, Vol. III.

Xu, Z.-T., Zhao, A.-D, Mei, Y. L., and Guo, Q.-S. (1979): Solar-Terrestrial Predictions Proceedings.

Zirker, J. B., editor (1977): Coronal Holes and High Speed Wind Streams, Colorado Associated University Press, Boulder, Colo., 454 pages.

#### References omitted:

Dodson, H. W. and Hedeman, E. R. (1972): Time variations in solar activity. in Solar-Terrestrial Physics/1970, Part I, E. R. Dyer, ed., D. Reidel, 151.

McIntosh, P. S. (1979): Annotated atlas of H-alpha synoptic charts for solar cycle 20(1964-1974). Upper Atmosphere Geophysics Report UAG-70, World Data Center A for Solar-Terrestrial Physics, Boulder, Colorado.

Roelof, E. C. and Krimigis, S. M. (1973): Analysis and synthesis of coronal and interplanetary energetic particle, plasma and magnetic field observations over three solar rotations. J. Geophys. Res., 78:5375.

Roelof, E. C. (1974): Coronal structure and the solar wind. in Solar Wind Three, C. T. Russell, ed., Inst. Geophys. and Planetary Phys., Univ. Calif. at Los Angeles, 98.



# A NEW METHOD OF VERY LONG-TERM PREDICTION OF SOLAR ACTIVITY

A. I. Ohl

Main Astronomical Observatory, Pulkovo, Leningrad, USSR

G. I. Ohl

Arctic and Antarctic Research Institute, Fentanka 34,  
Leningrad, USSR

Based on the close correlation between geomagnetic activity near the end of each solar cycle and Wolf's number in the maximum of the next 11-year cycle, a new method of very long-term prediction of solar activity is proposed. This method is applied to predict sunspot cycle No. 21.

We consider predictions of Wolf's numbers for the next 11-year cycle (or cycles) as very long-term predictions of solar activity. Existing methods of very long-term predictions (cf. Vitinsky, 1973) are based on either some internal regularities of solar activity (e.g., existence of many periods on the Wolf's numbers curve) or assumed planet influence on solar activity. The main shortcoming of these methods seems to be the absence of objective criteria for selecting the most effective periods or planets, since the results of such predictions depend considerably on this selection.

Furthermore, there is no physical basis to these methods, which may allow us to understand the physical nature of periods on the Wolf's numbers or planetary influence on solar activity. (Perturbation of the solar interior by planetary tidal forces may be insufficient.)

A method of solar activity prediction based on quite different assumptions has been developed during the last decade. Ohl (1966) has shown the existence of a very close statistical relationship between the minimum value of geomagnetic activity in each 11-year cycle and the height (i.e. Wolf's number at the maximum  $W_M$ ) of the next sunspot cycle. A linear regression equation was obtained,

$$W_M = 1.845M_{\min} + 66.3 \quad (r = +0.94), \quad (1)$$

where  $M_{\min}$  is the minimum average annual value of index  $M = 10(\sum K_p - 10)$  during a 11-year cycle. This equation was applied to the prediction of  $W_M$  in cycle 20. Taking  $M = 25$  for 1965, one can estimate  $W_M(20) = 112$ . This value differs a little from the observed value 106. Later Ohl (1968)

obtained a better correlation between  $W_M$  in the next sunspot cycle and index  $q$ , which was equal to the average value of the difference  $M' - W$  during the declining phase of the solar cycle, beginning when  $M'$  is greater than  $W$ , up to the minimum. Here  $M' = 8.6(\Sigma Kp - 10)$  is index  $M$  normalized to Wolf's numbers. This normalization was arranged so that the sum of the average annual values  $M'$  equaled the sum of  $W$  during all the years when the  $M$ -index was known. A close correlation exists between  $q$  and  $W_M$  in the next sunspot cycle with a correlation coefficient  $r = +0.95$ . Unfortunately, index  $q$  is inconvenient because to estimate it, one must know not only the epoch of the 11-year minimum, but values of Wolf's number and  $\Sigma Kp$  at this epoch.

In recent work (Ohl, 1976) a new index  $s$  was proposed. This new index is a sum of values  $M' - W$  during 3 years before the minimum of Wolf's numbers,

$$s = \sum_{i=-3}^{i=-1} (M'_i - W_i), \quad (2)$$

where  $M' = 8.6(\Sigma Kp - 10)$  and  $i = 0$  relates to the year of the minimum. The new regression equation for this index was

$$W_M = 0.694s + 45 \quad (r = +0.98). \quad (3)$$

The regression (3) was obtained for 9 solar cycles; statistical significance of this correlation coefficient is higher than 99.9%. See Figure 1.

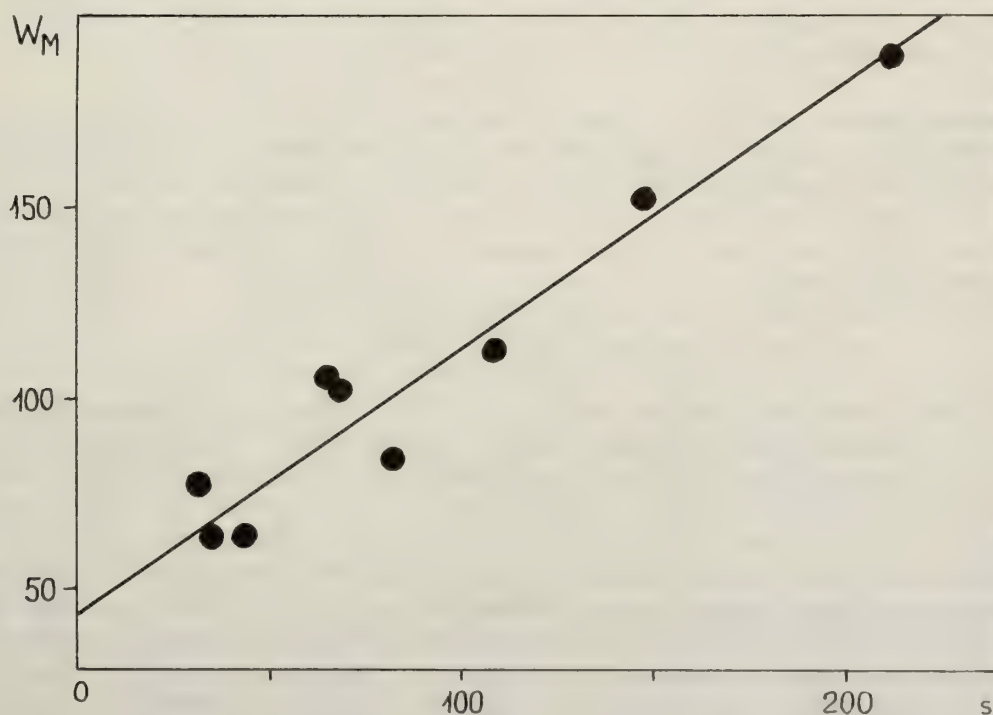


Figure 1. A relation between the indexes of recurrent magnetic disturbances and Wolf's number  $W_M$  in the maximum of the next 11-year cycle.

Table 1 contains values  $s^{(n)}$  and  $W_M^{(n)}$  for the same solar cycle and  $W_M^{(n+1)}$  for the next one.

TABLE 1

Sunspot cycle n number	$s^{(n)}$	$W_M^{(n)}$	$W_M^{(n+1)}$
11	35	139	64
12	83	64	85
13	44	85	64
14	68	64	104
15	31	104	78
16	109	78	114
17	147	114	152
18	212	152	190
19	66	190	106
20	198	106	

From Table 1, one more physically important conclusion stands out—values of index  $s^{(n)}$  are not connected statistically with  $W_M^{(n)}$ . The correlation coefficient is only +0.32, i.e. it differs slightly from 0.

Applying equation (3) to predict the height of cycle 21, we obtain  $W_M(21) = 183$ , i.e. the cycle starting in 1977 will be very high. It will be almost the same as cycle 19 with the maximum  $W_M(19) = 190$  in 1957. To determine the epoch of the maximum in cycle 21 we can use the regression between  $s$  and the duration  $t$  of the growth branch of the next cycle (Ohl, 1976). This regression is

$$t = -3.24 \log(s) + 10.2 \quad (r = -0.89). \quad (4)$$

It follows from this regression that  $t(21)$  will be equal to approximately 3 years, i.e. the maximum of cycle 21 will be in 1979. The method of prediction based on the equation (3) has been modified by Sargent (1978) and applied to smoothed monthly Wolf's numbers.

The physical meaning of index  $s$  lies in the fact that this index takes into account recurrent magnetic disturbances, occurring especially frequently on the declining branch of each 11-year cycle. A subtraction  $W_i$  from  $M_i$  eliminates the sporadic magnetic disturbances from the index  $s$ , since the number of sporadic disturbances is approximately proportional to the Wolf's numbers. Furthermore, it is known that recurrent magnetic disturbances correlate closely with coronal holes located above the Sun's unipolar magnetic regions (Ohl, 1971; Nolte et al., 1976; Sheeley et al., 1976; Zirker, 1977). Thus, the close relation between the recurrent magnetic disturbances



and the height of the next 11-year cycle and the complete absence of their relation with the height of the same solar cycle points to a physical connection between the extent of unipolar magnetic region (UMR) development on the decline branch of a given solar cycle with the development of bipolar magnetic sunspot regions in the next cycle. In other words, each cycle of solar activity begins some years before the minimum of Wolf's numbers, the first phase of a cycle being the development of UMR's. This phase ends shortly after the sunspot minimum when recurrent magnetic disturbances and UMR's practically disappear. Then the second phase of a solar cycle begins—the development of bipolar regions. During this phase, the increase of Wolf's numbers is observed; they reach their maximum and begin to fall. In the middle of the declining branch, UMR's occur again and recurrent disturbances appear, i.e. a new cycle of solar activity begins.

According to contemporary ideas (Babcock, 1961; Mullan, 1974; Yoshimura, 1975), bipolar groups result from the force tubes from the toroidal magnetic field rising through the solar surface. The toroidal magnetic field forms from the initial poloidal magnetic field because of the solar differential rotation. The poloidal field forms apparently not only from the rest of bipolar groups drifting to the poles of the sun, but it also originates from the convective zone (Stenflo, 1977).

Judging by the results of our work, one can assume that UMR's forming on the decline branch of the solar cycle are eruptions to the solar surface of the new poloidal magnetic flow, their number and intensity reflecting the strength of the new poloidal magnetic field development. It is natural to suppose also that the intensity of the toroidal magnetic field that appears at the second phase of the cycle (and the degree of bipolar regions development) is determined by the initial poloidal field formed during the first phase of the cycle. This explains the discovered physical relation between the intensity of bipolar regions at the second phase of the solar cycle and the degree of recurrent magnetic disturbances development at the first phase of the cycle.

Our results are obtained for nine sunspot cycles; therefore, one could doubt the truth of the obtained statistical relation. The auroral data are used for the testing of this relation. The data are taken from auroral catalogues for Sweden (Rubenson, 1882) and for Norway (Tromholt, 1902). These catalogues list the number of days with aurora observed in each year separately for some latitude zones. On the basis of these data, the mean cyclical curves of the number of aurora in a year (from 1721 to 1877) were constructed for each zone, the years of maxima of solar activity being considered as the key years. This procedure was applied to the even and odd cycles separately. On all mean curves, an appreciable variation of the number of aurora in a 11-year cycle was seen with the maximum number of aurora usually occurring near the maximum Wolf's number. On the curves for even cycles, constructed for the fourth zones of both catalogues (at the latitudes  $55^{\circ} - 61^{\circ}30'$ ), an increased number of aurora on the declining branch of the cycle was observed, which is typical for recurrent magnetic disturbances (Ohl, 1974). On the analogous curves for odd cycles, an increase was either not observed or was very weak. From this, one may conclude that the data for the fourth zones in Scandinavian catalogues describe to some extent the development of recurrent auroral phenomena.

A sum of annual numbers of aurora during the year of a solar minimum and the three years before it for the fourth zones of both catalogues were used as the index of recurrent phenomena. A correlation of this index ( $q_a$ ) with  $W_M$  in the next cycle was fairly high,  $r = +0.90$ . See Figure 2. The correlation of  $q_a$  with  $W_M$  in the same cycle is appreciably lower,  $r = +0.43$ . Excluding the anomalous cycle 7 whose declining branch was equal to only three years, we obtain that the correlation coefficient  $r(q_a, W_M^{(n+1)})$  increases to  $+0.96$ . In that cycle, recurrent auroral phenomena probably could not develop fully; therefore, we have every reason to exclude that cycle.

Thus, verification with absolutely independent data, including 13 cycles of solar activity (from cycles -2 to 10), confirms the correctness of our fundamental conclusion about a close correlation between recurrent magnetic disturbances and the height of the next 11-year cycle of Wolf's numbers.

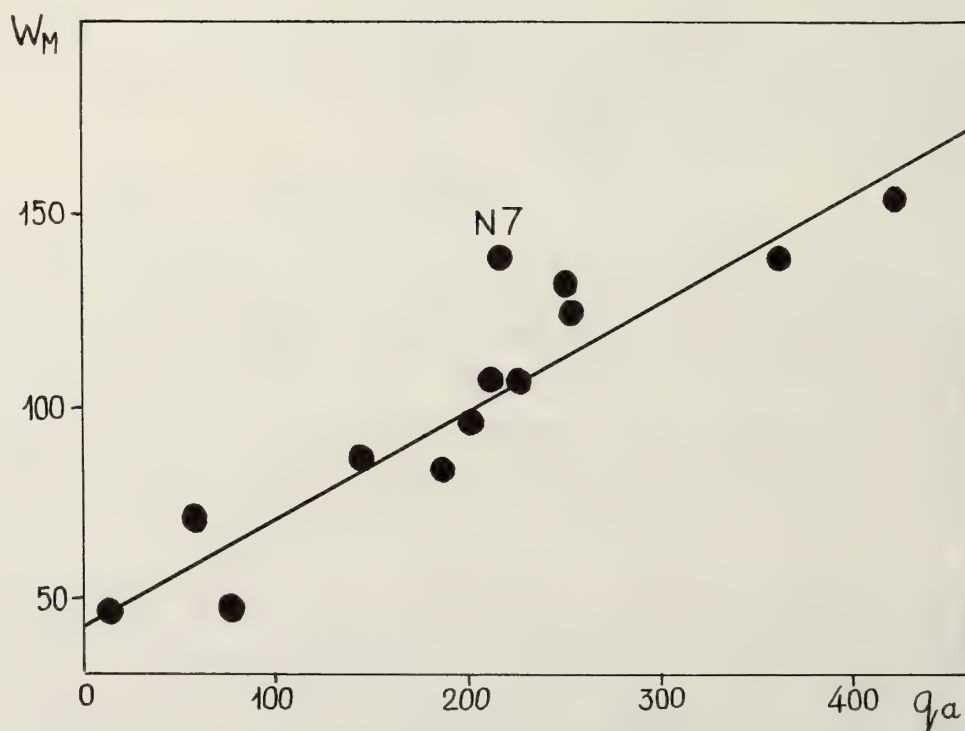


Figure 2. The relation between the index of aurora  $q_a$  and Wolf's number  $W_M$  at the maximum of the next 11-year cycle.

## REFERENCES

- Babcock, H. W. (1961): The topology of the Sun's magnetic field and the 22-year cycle. Astroph. Journ., 133:572.
- Mullan, D. J. (1974): Magnetic fields in the Sun. Journ. Frankl. Inst., 298:341.
- Nolte, J. T., A. S. Krieger, A. F. Timothy, R. E. Gold, E. C. Roelof, G. Vaiana, A. J. Lazarus, J. D. Sullivan, and P. S. McIntosh (1976): Coronal holes as sources of solar wind. Solar Phys., 46:303.
- Ohl, A. I. (1966): Wolf's number prediction for the maximum of the cycle 20 (in Russian). Solar Data, 12:84.
- Ohl, A. I. (1968): Wolf's number prediction for the maximum of the current solar cycle (N 20) (in Russian). Probl. of Arctic and Antarctic, 28:137.
- Ohl, A. I. (1971): On the physical nature of the 11-year variation of geomagnetic disturbances (in Russian). Geomagn. and Aeronomy, 11:647.
- Ohl, A. I. (1974): The cyclical variation of auroral phenomena (in Russian). In: High-latitude geophysical phenomena, Leningrad, "Nauka", 7-22.
- Ohl, A. I. (1976): The preliminary prediction of some parameters of 21 solar cycle (in Russian). Solar Data, 9:73.
- Rubenson, R. (1882): Catalogue des aurores boréales observée en Suède (1536-1877). Kongl. Svenska Vet. Akad. Handl., 18(1): 1-300.
- Sargent, H. H. (1978): A prediction for the next sunspot cycle. Preprint.
- Sheeley, N. R., J. W. Harvey, and W. C. Feldman (1976): Coronal holes, solar wind streams, and recurrent geomagnetic disturbances: 1973-1976. Solar Phys., 49:271.
- Stenflo, J. O. (1977): Solar-cycle variations in the differential rotation of solar magnetic fields. Astron. and Astrophys., 61:797.
- Tromholt, S. (1902): Catalog der in Norwegen bis Juni 1878 beobachteten Nordlichter. Kristiania, J. Dybwad, 422 pp.
- Vitinsky, Ju. I. (1973): Cycles and predictions of solar activity (in Russian). Leningrad, "Nauka", 257 pp.
- Yoshimura, H. (1975): A model of the solar cycle driven by the dynamo action of the global convection in the solar convection zone. Astroph. Journ. Suppl. Ser., 29:467.
- Zirker, J. B. (1977): Coronal holes and high-speed wind streams. Rev. Geoph. Space Phys., 15:257.



# NEW METHODS FOR PREDICTING THE MAGNITUDE OF SUNSPOT MAXIMUM

G. M. BROWN

Department of Physics, University College of Wales,  
Aberystwyth, UK

Three new and independent methods of predicting the magnitude of a forthcoming sunspot maximum are suggested. The longest lead-time is given by the first method, which is based on a terrestrial parameter measured during the declining phase of the preceding cycle. The second method, with only a slightly shorter foreknowledge, is based on an interplanetary parameter derived around the commencement of the cycle in question (sunspot minimum). The third method, giving the shortest prediction lead-time, is based entirely on solar parameters measured during the initial progress of the cycle in question. Application of all three methods to forecast the magnitude of the next maximum (cycle 21) agree in predicting that it is likely to be very similar to that of cycle 18.

## 1. Introduction

It is well known that the magnitude of the maxima of solar activity in the eleven-year cycle varies in an apparently random manner from cycle to cycle. In view of the importance of this parameter in relation to a wide range of solar-terrestrial phenomena, a reliable method of prediction would have considerable practical significance. The literature contains many examples of such efforts, but so far the success rate has been unspectacular.

Methods of prediction of the size of sunspot maximum fall into four broad categories. In decreasing order of lead-time these are based on:

- (i) long-period recurrence tendencies in previous cycles, or statistical analyses of the distribution of peak sizes ;
- (ii) behaviour of the preceding cycle, particularly its declining phase ;
- (iii) behaviour at solar minimum, considered as the start of the cycle in question ;
- (iv) initial progress of the current cycle.

In the absence of a detailed physical understanding of the cause of the solar cycle, all prediction methods are necessarily empirical, and it is therefore not surprising to find a wide range of values for a given cycle. For example, estimates of the maximum annual mean sunspot number ( $R_a$ -max) for the forthcoming cycle number 21 vary from 37-67 (Bonov, 1970) to 195 (Ramaswamy, 1977). When comparing magnitudes it is important to distinguish between  $R_a$ -max and the (invariably larger) maximum smoothed monthly mean sunspot number,  $R_m$ -max, which is often considered. Over the last 13 cycles the average difference between these two measures has been about 4 per cent, but it has ranged from 0.1% (cycle 18) to 17.1% (cycle 12).

In the present paper an outline is given of three new possible methods of prediction, one based on each of the approaches (ii), (iii) and (iv) above. The first two of these, being dependent on the approach to minimum in a previous cycle or conditions obtaining at the minimum, are best considered in relation to  $R_a$ -max; the third method considers the initial development of the cycle in question in terms of monthly means and leads to an estimate of  $R_m$ -max.

## 2. AQD prediction method

It is well known that the 27-day recurrent type of geomagnetic storm occurs preferentially during the declining phase of a solar cycle, and Ohl (1971) found that the magnitude of this type of magnetic activity on the declining part of a cycle is directly related to the maximum value of sunspot number in the next cycle. The relation has been confirmed by Sargent (1977) and developed into a prediction technique holding considerable promise. The current 'Solar Geophysical Data' predictions published by the U.S. Department of Commerce (NOAA) include the use of the so-called "Ohl-Sargent method". Another recent proposal that the course of a solar cycle is somehow linked with the behaviour of the preceding cycle has been given by Ramaswamy (1977), who finds that the relative skewness of one cycle is linearly related to the maximum annual mean sunspot number of the next cycle, expressed as a ratio to the maximum of the current cycle.

The first indication that a solar-controlled terrestrial parameter might exhibit a half-cycle foreknowledge of the course of a cycle was introduced by Brown and Williams (1969) in considerations of the variability of the quiet-day geomagnetic field. This work was developed into a new solar-terrestrial relationship (Brown, 1974) which held good for eight solar cycles and had distinct prediction possibilities. The basis of this method will now be described briefly.

Both the amplitude and phase of the quiet-day solar diurnal variation of the horizontal component ( $H$ ) of the earth's magnetic field,  $Sq(H)$ , vary considerably from day to day. Here we are concerned only with the phase variability at a mid-latitude northern hemisphere station. The distribution of occurrence of the time of the daily minimum of  $H$  at Greenwich, or its



successor observatories Abinger and Hartland, has been studied for all the International Quiet Days over the period 1884-1972. Although the most usual time of H-min is an hour or so before noon, there is in practice a very wide variation in this time. Reference should be made to Figure 1 in Brown (1975) for details. On the basis of this study each day may be classified as either a 'normal quiet day' (NQD) or an 'abnormal quiet day' (AQD). The choice of the dividing line between these categories is not crucial to the subsequent development of the method and has been made arbitrarily.

It has been found that the incidence of AQDs exhibits a marked solar control, being greatest in local winter and near sunspot minimum epochs. This latter property has been investigated in detail (Brown, 1974) and shown to exhibit an apparent and remarkable ability to anticipate the magnitude of the forthcoming sunspot maximum in each cycle. Thus, the variation in the annual number of AQDs displays an inverse solar cycle such that the increase in AQD-count ( $Q_A$ ) from its minimum near sunspot maximum to its maximum near the following sunspot minimum varies in sympathy with the increase in sunspot number from this minimum to the succeeding maximum. Similarly, the decrease in  $Q_A$  over the increasing phase of the solar cycle matches the decrease in sunspot number over the declining phase following. The original analysis was based on three-yearly running means of  $Q_A$  and of annual mean sunspot number  $R_a$ , but the latter is unsuitable for prediction applications since the maximum value of  $\bar{R}_a$  can differ significantly from  $R_a$ -max. Furthermore, for prediction of the magnitude of the next solar maximum it is obviously essential to confine attention to the increasing part of the  $Q_A$  cycle (i.e. from sunspot maximum to minimum).

These facts have been taken into account in Figure 1 which shows the variation from 1885 to 1977 of  $R_a$  (curve) and of the measured increase in three-yearly running mean AQD-count ( $\Delta \bar{Q}_A$ ) over each declining sunspot half-cycle (asterisks plotted at the year of maximum value of  $\bar{Q}_A$ ). The error bars on the values of  $\Delta \bar{Q}_A$  are drawn at  $\pm 20\%$  to indicate a reasonable estimate of the order of accuracy of this measurement. It is seen that, within this range, each value of  $\Delta \bar{Q}_A$  successfully maps the magnitude of the forthcoming  $R_a$ -max. Figure 2 shows that there is a reasonably linear relation between these two parameters. Assuming that the relation holds for future cycles, measurement of  $\Delta \bar{Q}_A$  over the period of declining activity of a cycle may be directly used to forecast the peak annual mean sunspot number of the following cycle from Figure 2. In an individual case such a predicted value will be subject to an uncertainty of about  $\pm 20\%$ , owing to the fact that  $\bar{Q}_A$  is a relatively small number.

### 3. IMF prediction method

Perhaps the first suggestion that the development of a sunspot cycle proceeds to its maximum in a predetermined way from its commencement at



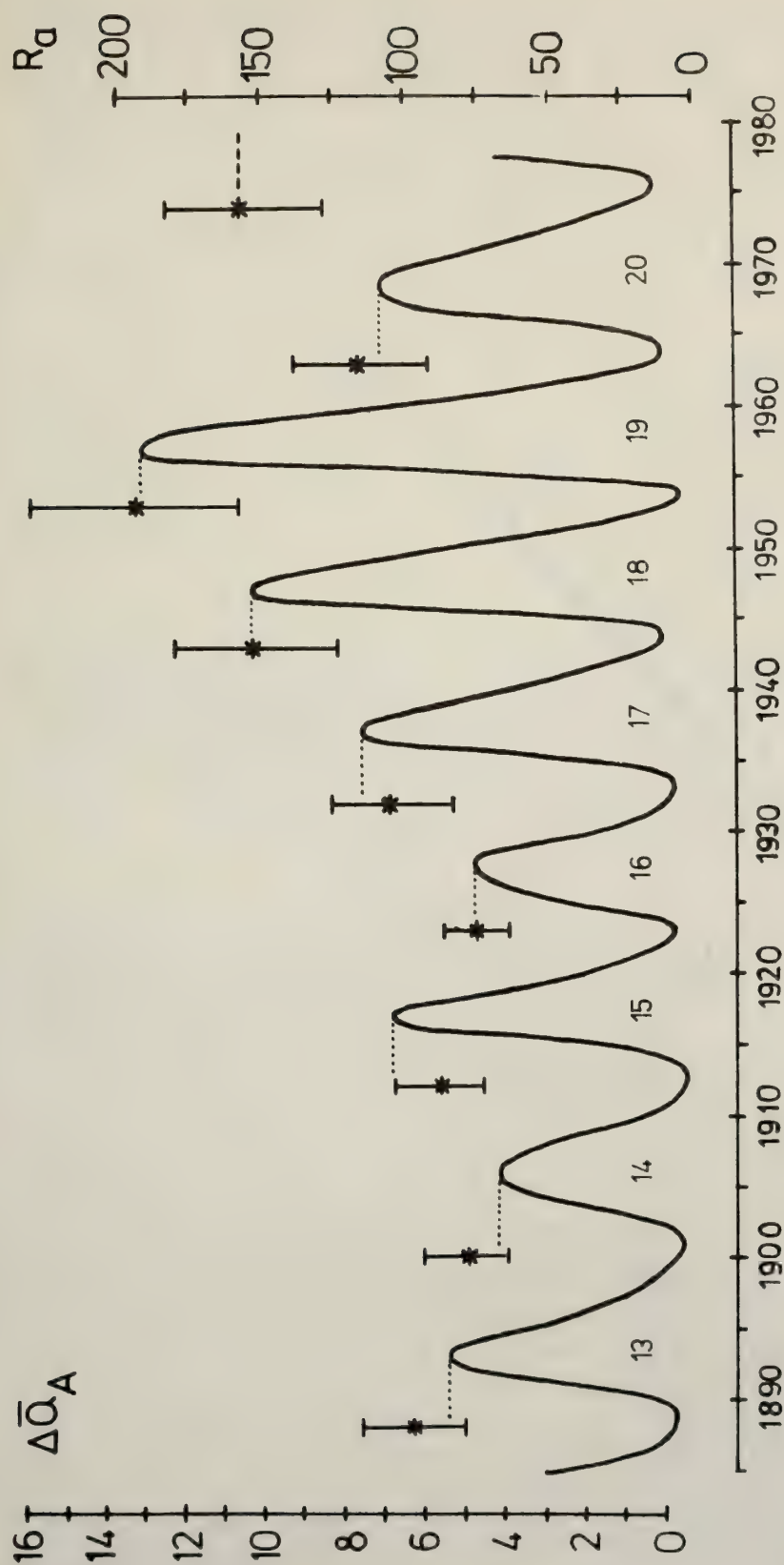


FIG. 1 VARIATIONS OF ANNUAL MEAN SUNSPOT NUMBER  $R_a$  (CURVE) FOR CYCLES 13 TO 20, AND INCREASE IN THREE-YEARLY RUNNING MEAN AQD-COUNT  $\Delta \bar{Q}_A$  OVER DECLINING PHASES (ASTERISKS).

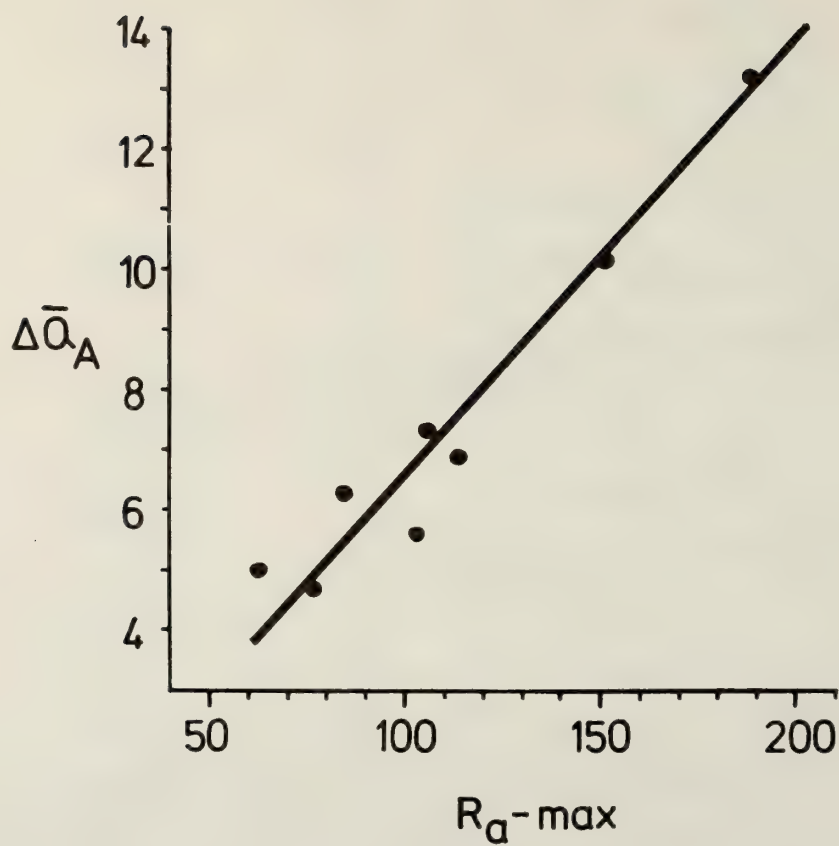


FIG. 2 RELATION BETWEEN  $\Delta\bar{Q}_A$  AND MAXIMUM ANNUAL MEAN SUNSPOT NUMBER  $R_a\text{-max}$  FOR CYCLES 13 TO 20.

sunspot minimum was given by Waldmeier (1939), who showed that there was a correlation between the average smoothed heliographic latitude at which the spots of a new cycle develop and the subsequent maximum activity reached. This result was based on a smoothed analysis of seven cycles, and is not directly applicable for prediction purposes to an individual cycle. Another example of the principle that 'the minimum determines the maximum' drawn from the really long-term (secular) behaviour of sunspot number itself has been given by Brown (1976). This also is not amenable to prediction applications, since it is based on long period (80-year?) variations extracted by smoothing the data over several solar cycles.

As a new example of an apparent relation between the behaviour of a solar parameter at sunspot minimum and the magnitude of the subsequent sunspot maximum, which might lend itself to development into a prediction technique, we turn to measurements of the sector structure of the interplanetary magnetic field. In particular, the number of 'A-days', when the IMF is directed away from the sun, is a parameter which suggests itself since it maximises at sunspot minimum. Moreover, Wilcox and Ness (1965) have explicitly identified the source of A-sectors lasting ten days or more with a ghost unipolar magnetic region (GUMR) on the sun which, together with its associated UMR, may well be linked with the solar 'M regions' responsible for recurrent magnetic storms (Babcock and Babcock, 1955; Bumba and Howard, 1966; Sakurai, 1966) mentioned at the beginning of the previous section. It was therefore decided to look at the distribution of runs of 'A-days', adopting a criterion of a minimum of five consecutive days of away-polarity as a meaningful sequence of consistent IMF direction.

In order to study the variation over as many solar cycles as possible it was necessary to use the tables of daily inferred IMF polarity back to 1926 given by Svalgaard (1972). These are derived from examination of the type of diurnal variation of the geomagnetic field within the polar cap, and are known to be inaccurate on certain occasions. Russell and Rosenberg (1974) determined that the maximum accuracy attainable in this method is about 88%; in practice it is usually less than this. Russell, Fleming and Fougere (1975) showed that for the more recent data the success rate in inferring A-days was some 5% less than that attained for 'toward' days ('C' days in Svalgaard's original notation) but that "for many statistical studies the present-day inferences of interplanetary polarity are adequate substitutes for in situ observations".

The solid curve in Figure 3 shows the variation over the period  $\pm 3$  years from sunspot minimum (indicated by the downwards-pointing arrow on the upper abscissa scale) of the total number per year of consecutive A-days equal to or exceeding 5 in any one sequence. The evident tendency for a maximum occurrence of 'long-lived' A-sectors at sunspot minimum is probably another aspect of the established solar cycle variation in the number of A-days mentioned above. The dotted curve shows the annual mean sunspot number over the period -2 to +3 years from sunspot maximum (indicated by



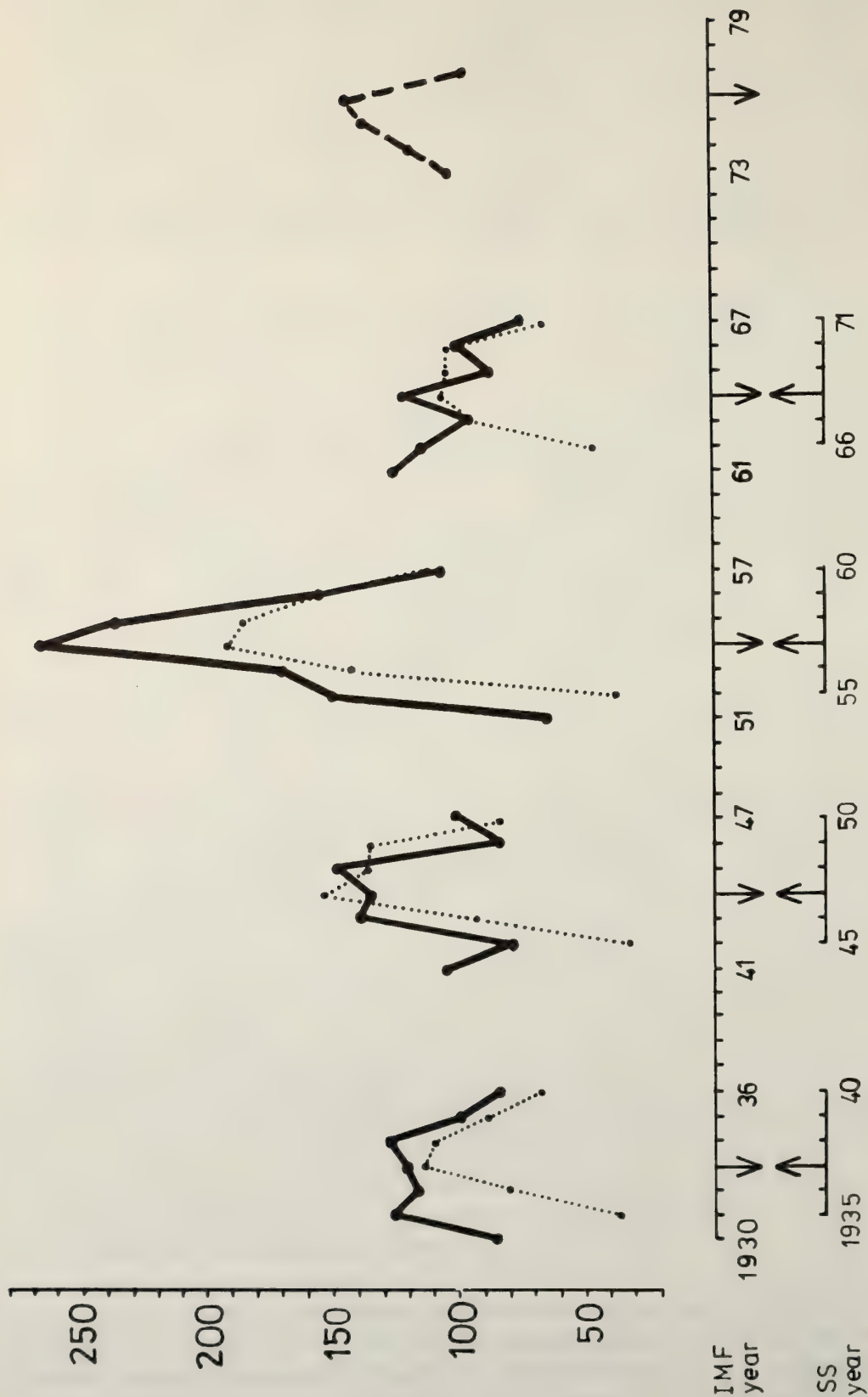


FIG. 3 VARIATIONS OF OCCURRENCE OF LONG-LIVED A-SECTORS OF THE IMF (SEE TEXT) AROUND SUNSPOT MINIMUM (SOLID CURVE; UPPER YEAR SCALE) AND ANNUAL MEAN SUNSPOT NUMBER AROUND SUBSEQUENT SUNSPOT MAXIMUM (DOTTED CURVE; LOWER YEAR SCALE).

DOWNWARD ARROWS: SUNSPOT MINIMUM; UPWARD ARROWS: SUNSPOT MAXIMUM.

the upward-pointing arrow on the lower abscissa scale) where the year of each maximum has been synchronised to that of the preceding minimum. Conveniently, the ordinate scales for the two parameters are the same. The data cover solar cycles 17 to 20. Although there is no precise quantitative agreement between the magnitudes, there is clearly an approximate relationship over these four cycles between the size of  $R_a$ -max and that of preceding occurrence of long-lived A-sectors at sunspot minimum. Of particular interest is the extra large forerunner to cycle 19, followed by the marked reduction for cycle 20.

In this form, the result of Figure 3 may be useful to predict an order of magnitude of the forthcoming sunspot maximum from measurements of the polarity of the IMF at the preceding minimum. A refinement of the method may enable a quantitative prediction to be made.

#### 4. SVF prediction method

Probably the most-used solar activity prediction methods hitherto have been based on empirical relationships between the rate of rise of activity (usually measured by sunspot number) in the early stages of a solar cycle and the maximum ultimately reached. In most cases these techniques involve estimating the total rise time ( $t_r$ ) of a cycle, which is inversely related to  $R_m$ -max, and therefore in principle they predict the time of occurrence as well as the magnitude of the next maximum. However, no one accepted relation between  $t_r$  and  $R_m$ -max has been uncovered, and the present diversity of relations is reflected in a wide range of predictions for a given cycle.

It has been known for a long time that, in many ways, photospheric faculae have a more fundamental nature than sunspots. Faculae have a much longer lifetime than spots, and they often provide a connecting link between successive spot groups which may revive in a facular region after becoming invisible. Their duration also endows them with a greater ability than spots to preserve a record of activity on the invisible hemisphere of the sun. Such considerations may well give significance to the so-called 'areas index' of solar activity introduced by Xanthakis (1969) which combines both sunspot and faculae areas into a new index. In terms of the initial development of a solar cycle as a precursor of its future magnitude, it appears relevant to investigate the rate at which sunspot areas develop relative to faculae areas. Accordingly, we have considered the relation between 13-month running mean sunspot areas  $A_S$  and corresponding faculae areas  $A_F$  for the first few years of each solar cycle since 1874, using the Greenwich photoheliographic results, corrected for foreshortening. The results of this investigation including consideration of the influence on the data of the observational impossibility of observing faculae in the central part of the solar disc, will be published elsewhere (Brown and Evans, 1980).



In general, the 'spots versus faculae' (SVF) plots show a sigmoid type of pattern over the growing part of a cycle, which can become rather ill-defined towards solar maximum, but over the first two years or so the linear correlation between the two areas is very high (0.97-1.00) for all the cycles 12 to 20. The slope ( $m_{1500}$ ) of the least-squares line fitted to the SVF data, taken arbitrarily as far as  $A_F = 1500$  millionths, for each cycle has been found to change systematically with the magnitude of the cycle. In fact, there is an almost one-to-one relation between  $m_{1500}$  and  $R_m$ -max from cycle to cycle, as is clear from Figure 4 where the solid thick line shows the variation in  $m_{1500}$  and the dotted line that in  $R_m$ -max. Figure 5(a) shows that these two parameters are approximately linearly related when  $R_m$ -max is plotted on a logarithmic scale.

An alternative presentation of the above result, which is simpler to derive and seems to yield an even better-defined relationship, is shown in the solid fine line in Figure 4. The ordinate here  $(A_S)_{1500}$  is the value of  $A_S$  on the rising part of each cycle for  $A_F = 1500$ , estimated by simple linear interpolation between the two nearest running mean monthly values which straddle this value of  $A_F$ . This parameter is clearly related to  $m_{1500}$ , but it does not concern itself with the detail of the actual progression to the chosen cut-off value of  $A_F$  (which in practice proceeds rather differently in individual cycles). The way in which  $(A_S)_{1500}$  varies in sympathy with the subsequent amplitude of the sunspot cycle to which it belongs (dotted line in Figure 4) is rather remarkable. Figure 5(b) shows that there is a very good linear relation between  $(A_S)_{1500}$  and  $R_m$ -max.

In practice, for the nine cycles considered the arbitrarily selected threshold value of  $A_F = 1500$  is reached after an average of 23 months from the beginning of the cycle. Allowing for the 13-month smoothing requirement, this means that the appropriate parameter,  $m_{1500}$  or  $(A_S)_{1500}$  as defined above, should be available within the first  $2\frac{1}{2}$  years of a cycle, which gives a useful lead on the sunspot maximum occurring usually  $1\frac{1}{2}$  to  $2\frac{1}{2}$  years later. From the graphs of Figure 5 the predicted value of  $R_m$ -max may be inferred. The r.m.s. deviation of the points from the lines in these figures is only 5 or 6 in sunspot number; if the suggested relation holds in the future it implies a possibility of predicting the magnitude of a forthcoming sunspot maximum to better than 10%. Investigation of the SVF relation taken to other cut-off values of  $A_F$  may well show that the prediction lead-time can be increased.

## 5. Applications to cycle 21

Gleissberg (1971) lists nine different predictions for the magnitude of the forthcoming sunspot maximum, which range from 30 to 145 for  $R_m$ -max. More recently, several other estimates have been published, with a general tendency favouring higher values. The degree of confusion is well illustrated



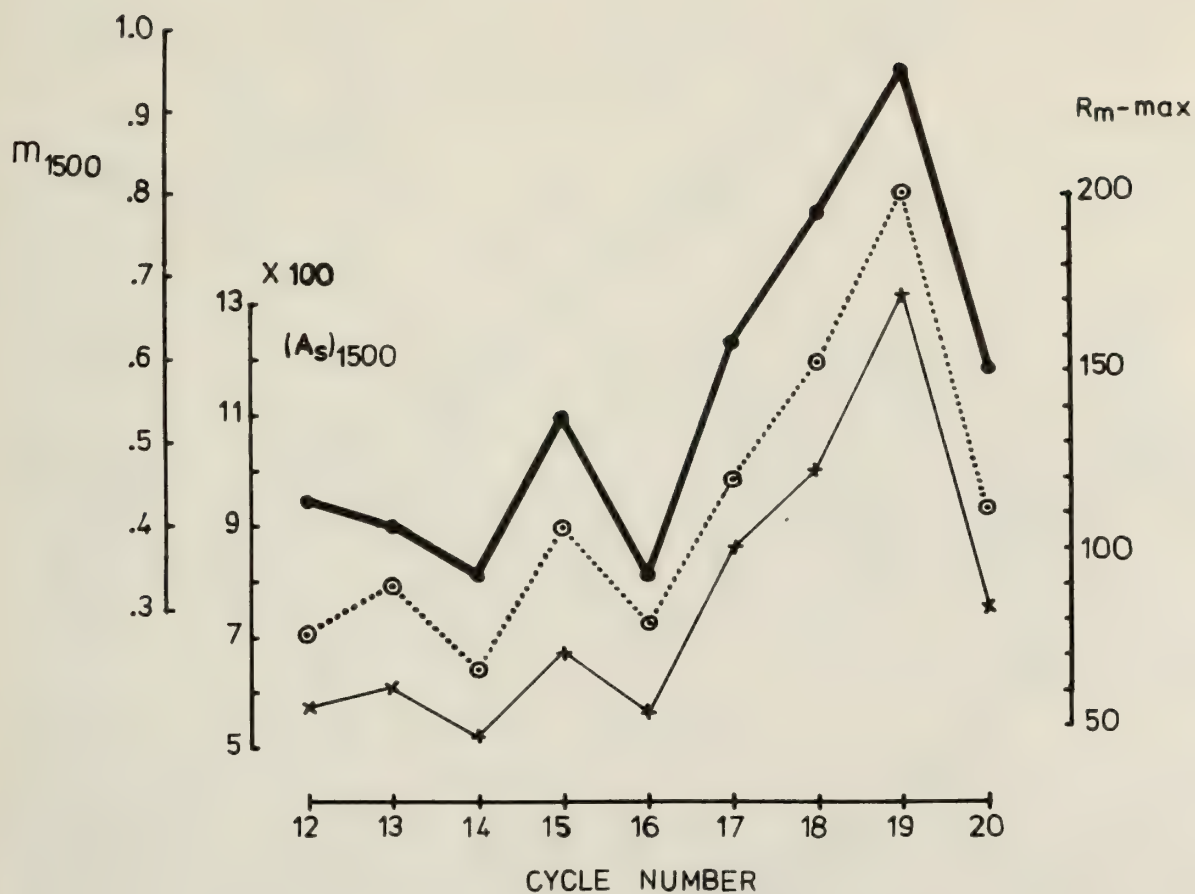


FIG.4 COMPARISON OF PARAMETERS  $m_{1500}$  (THICK LINE) AND  $(A_S)_{1500}$  (THIN LINE) IN THE SVF RELATIONSHIP WITH PEAK SMOOTHED MONTHLY MEAN SUNSPOT NUMBER  $R_m - \max$  (DOTTED LINE) FOR CYCLES 12 TO 20.

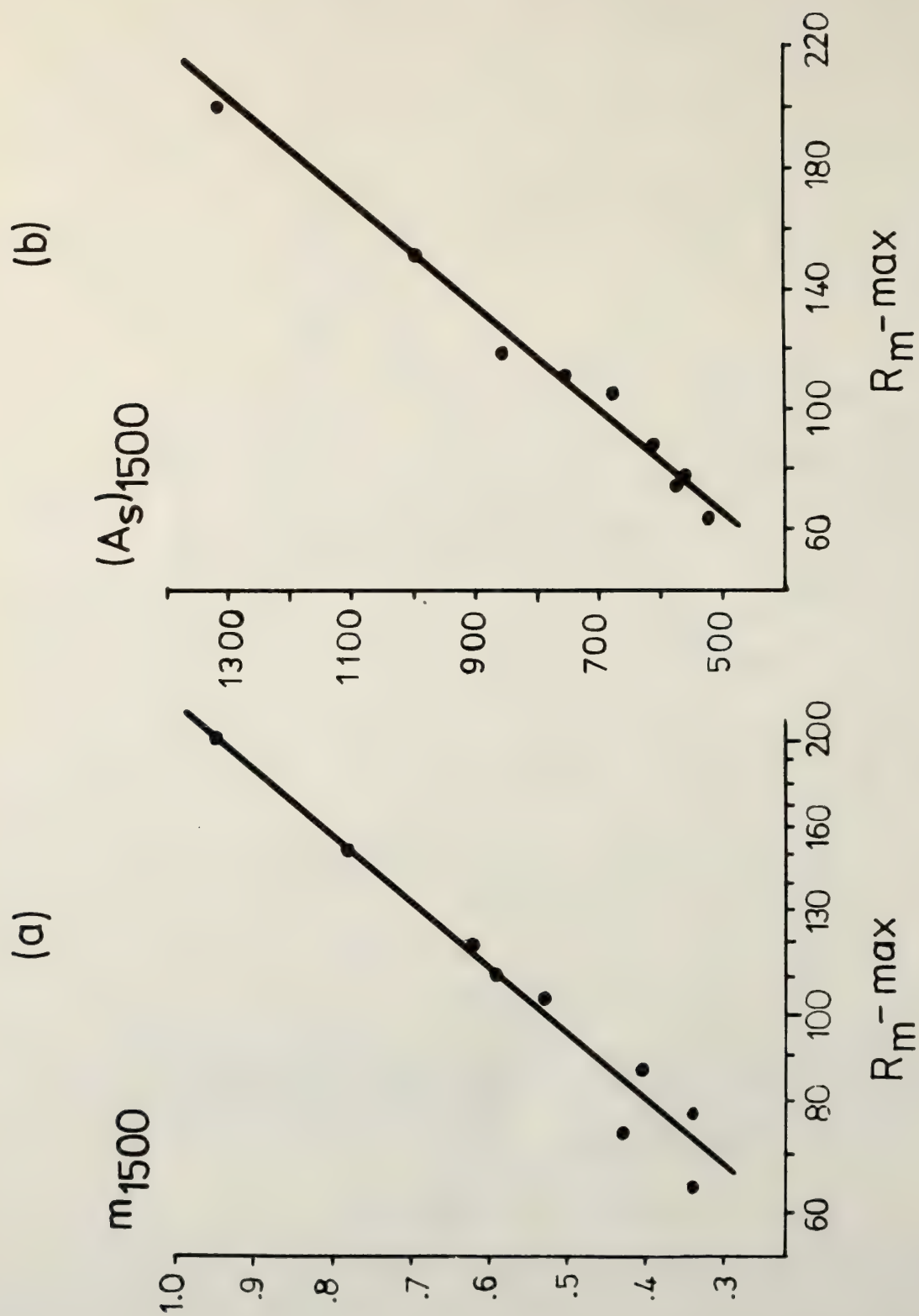


FIG. 5 RELATION BETWEEN (a)  $m_{1500}$  AND  $R_m - \max$ , AND (b)  $(A_S)_{1500}$  AND  $R_m - \max$  FOR CYCLES 12 TO 20.

by quoting two of the most recent predictions of  $R_a$ -max: Ramaswamy (1977) proposes a value of 195, while Xanthakis and Poulakos (1978) propose 82. Current indications are that activity has risen fast and may well go to a high peak.

Within the context of the AQD method discussed in Section 2, the value of  $\Delta \bar{Q}_A$  for Hartland magnetic data for the declining part of cycle 20 is 10.5, with the peak value of  $\bar{Q}_A$  in 1974. This value is included on Figure 1, and is close to the 10.2 for cycle 18. From Figure 2 it leads to a predicted value of  $155 \pm 31$  for  $R_a$ -max (cf. 151.6 for cycle 18).

As mentioned before, the IMF method as discussed in Section 3 does not lead to a quantitative prediction. However, on Figure 3 the dashed line gives an estimate of the likely variation of the annual number of 'long-lived' A-sectors, as previously defined, over the period of the sunspot minimum of 1976. These values have been obtained from the U.S. (NOAA) 'Solar Geophysical Data' tables of inferred IMF polarities, which are not entirely compatible with Svalgaard's data used in the preceding four cycles since they include two estimates of IMF direction per day (one for each polar cap) and also introduce additional categories of 'indeterminate' and 'missing' information. In this analysis, missing days have been given the benefit of the doubt when they occurred in a sequence of A-days, and one half of the maximum number of additional A-days arising had the indeterminate days been taken as A-days have been included. It is not established that this approach really gives an adequate 'Svalgaard estimate' of the occurrence of A-days, but it seems reasonable. With these reservations, the inference from Figure 3 is that cycle 21 is likely to resemble cycle 18 in magnitude, in agreement with the AQD prediction method.

Regrettably, it is not possible to obtain for the beginning of cycle 21 a precise figure for either  $m_{1500}$  or  $(A_S)_{1500}$  in the notation of the SVF method discussed in Section 4. This is because the Royal Greenwich Observatory data for 1976 are not yet available and, furthermore, it has been decided to discontinue the long series of daily measurements of sunspot and faculae areas from the end of 1976. From January 1977 area data for one day a week only are available for Greenwich, and these allow a cruder estimate of the monthly mean values of  $A_S$  and  $A_F$  to be obtained. A change in scaling practice has also been introduced, in that a significant number of polar faculae are now included in the weekly measurements, whereas these were essentially excluded in the previous tabulations. In an effort to preserve compatibility with the latter, all polar faculae in the weekly listings have been ignored in the calculation of the daily total  $A_F$ . The complete absence of data for 1976 makes it impossible even to estimate a slope  $m_{1500}$  for the SVF plot, but the 1977-78 limited data lead to an estimate of  $992 \pm 130$  for  $(A_S)_{1500}$ . From Figure 5(b) this transcribes into a predicted value of  $151 \pm 24$  for  $R_m$ -max, almost identical to the value 151.8 for cycle 18.

It is interesting (even compelling?) to find that all three of these new



and totally independent methods of prediction agree in forecasting that cycle 21 will be of very similar magnitude to that of cycle 18 with a peak sunspot number of about 152. This is also in agreement with the NOAA 'Solar Geophysical Data' prediction based on the Ohl-Sargent method mentioned at the beginning of Section 2, the latest estimate (August 1978) being  $153 \pm 35$ .

## 6. Discussion

In this paper we have introduced three completely new possible methods of predicting the magnitude of a forthcoming sunspot maximum. One (AQD method) is based entirely on a terrestrial parameter (quiet day magnetic field) measured during the declining phase of the previous cycle; one (IMF method) is based on an interplanetary parameter (polarity of the IMF) measured or inferred at the time of sunspot minimum; one (SVF method) is based entirely on solar parameters (sunspot and faculae areas) measured during the first year or two of the cycle in question. Like all other prediction techniques in this field, the methods are entirely empirical with no known physical basis. Running through them all is an inherent assumption that the magnitude of a solar maximum is predetermined, in an as yet unknown way, at an earlier stage of development which spans on either side of the very beginning of the cycle at sunspot minimum. Waldmeier's (1939) law regarding the latitude of first appearance of the spots of a new cycle contains the germ of the concept that the 'minimum determines the maximum'. The fact that the solar cycle is, in practice, a complicated sequence of events with different phase relationships further confuses the interpretation. Despite the lack of a detailed physical understanding there is a degree of logic in the concepts introduced here. By way of conclusion some general points may be made.

Basically, active regions result from the appearance of magnetic fields at the solar surface and they are evident visually as photospheric faculae. Sunspots may or may not appear in these regions for a part of their lifetime. Since solar activity in its broadest sense may be regarded as some kind of solar magnetic field oscillation, it seems very pertinent in studying the growth of a cycle (albeit measured in terms of sunspot development) to consider the relative extents of the spot and faculae areas. In an important report from the Royal Greenwich Observatory (1923) it was noted that the proportion of faculae areas unassociated with spots is highest at sunspot minimum (averaging 30%) and falls to only 6% at sunspot maximum. It does not seem unreasonable to suggest that the more intense is a cycle of activity the less is this proportion, i.e. the greater is the development of spots within the faculae areas, during the early stages of the cycle. This would give physical meaning to the SVF results.

An important line of communication between the sun and the earth is the interplanetary magnetic field, which, according to Abdel-Wahab and

Goned (1974), has a particularly stable sector structure at sunspot minimum. The suggestion from the IMF results presented here is that the occurrence of long-lived A-sectors at this solar epoch is an important feature predetermining the magnitude of the subsequent cycle. Bumba and Howard (1965) pointed out that magnetically active regions are concentrated into 'complexes of activity' near sunspot minimum, in the largest of which UMRs form. These and their accompanying GUMRs, which have been specifically linked with A-sectors of the IMF by Wilcox and Ness (1965), are associated with open magnetic field lines and the development of coronal holes at the coronal level. In their turn, coronal holes are believed to be the source of high speed solar wind streams that cause recurrent magnetic storms. The use of recurrent magnetic activity as a precursor of solar activity (Ohl, 1971) is thus consistent with this picture.

Finally, it may be noted that the IMF certainly influences the diurnal variation of the earth's magnetic field at high latitudes (this being the basis of Svalgaard's (1972) method of inferring IMF polarity from geomagnetic measurements) and there have been suggestions of an influence at lower latitudes too. Thus Kane (1971) proposed that during quiet days the magnitude of the IMF partly controls the day-to-day variability of Sq range at low latitudes, and Matsushita et al (1973, 1975) observed that the position of the Sq focus (at geomagnetic latitude  $20^{\circ}$  -  $40^{\circ}$ ) varied with the IMF sector polarity. Such observations may be relevant in attempting a physical explanation of the AQD results presented here, which imply that the phase variability of Sq at a midlatitude station is sensitive to the level of solar activity half a cycle ahead.

#### Acknowledgements

The author wishes to thank Mr. P. J. Rudd of the Solar Activity Service at the Royal Greenwich Observatory, Herstmonceux Castle, and Mr. D. R. Evans of the UCW Physics Department, Aberystwyth, for helpful discussions and assistance in acquiring and analysing, respectively, the data used for the work described in section 4.

#### References

- Abdel-Wahab, S., and A. Goned (1974) : Solar cycle dependence of periodic variations in geomagnetic Kp-index. Planet. Space Sci., 22 : 537.
- Babcock, H. W., and H. D. Badcock (1955) : The sun's magnetic field, 1952-1954. Astrophys. J., 121 : 349.
- Bonov, A. D. (1970) : Soln. Dann. No. 7, 111.
- Brown, G. M. (1974) : A new solar-terrestrial relationship. Nature, 251 : 592.



- Brown, G.M. (1975) : Sq variability and aeronomic structure. J. Atmos. Terr. Phys., 37 : 107.
- Brown, G.M. (1976) : What determines sunspot maximum? Mon. Not. Roy. Astron. Soc., 174 : 185.
- Brown, G.M. and D.R. Evans (1980) : in preparation.
- Brown, G.M. and W.R. Williams (1969) : Some properties of the day-to-day variability of Sq(H). Planet. Space Sci., 17 : 455.
- Bumba, V. and R. Howard (1965) : Large-scale distribution of solar magnetic fields. Astrophys. J., 141 : 1502.
- Bumba, V. and R. Howard (1966) : A note on the identification of "M" regions. Astrophys. J., 143 : 592.
- Gleissberg, W. (1971) : The probable behaviour of sunspot cycle 21. Solar Phys., 21 : 240.
- Kane, K.P. (1971) : Relationship between day to day variability of Sq and interplanetary plasma parameters. Nature Phys. Sci., 232 : 21.
- Matsushita, S. (1975) : IMF polarity effects on the Sq current focus location. J. Geophys. Res., 80 : 4751.
- Matsushita, S., J.D. Tarpley and W.H. Campbell (1973) : IMF sector structure effects on the quiet geomagnetic field. Radio Sci., 8 : 963.
- Ohl, A.I. (1971) : Physics of the 11-year variation of magnetic disturbances. Geomag. Aeron. (English translation), 11 : 549.
- Ramaswamy, G (1977) : Sunspot cycles and solar activity forecasting. Nature, 265 : 713.
- Royal Greenwich Observatory (1923) : Solar faculae - their occurrence and distribution in latitude. Mon. Not. Roy. Astron. Soc., 84 : 96.
- Russell, C.T., B.K. Fleming and P.F. Fougere (1975) : On the sensitivity of modern inferences of interplanetary magnetic polarity to geomagnetic activity. J. Geophys. Res., 80 : 4747.
- Russell, C.T. and R.L. Rosenberg (1974) : On the limitations of geomagnetic measures of interplanetary magnetic polarity. Solar Phys., 37 : 251.
- Sakurai, K. (1966) : On the large-scale structure of the sun's magnetic fields during the declining and minimum phases of the solar activity cycles. Rep. Ionosph. Res. Japan, 20 : 501.
- Sargent, H.H. (1977) : A prediction of the next sunspot maximum. Trans. Amer. Geophys. Union, 58 : 1220 (abstract only).



- Svalgaard, L. (1972) : Interplanetary magnetic sector structure 1926 - 1971.  
Geophys. Papers R-29, Danish Meteorological Institute,  
Charlottenlund.
- Waldmeier, M. (1939) : Astr. Mitt. Zurich, 14 : 439.
- Wilcox, J. M. and N. F. Ness (1965) : Quasi-stationary corotating structure  
in the interplanetary medium. J. Geophys. Res., 70 : 5793.
- Xanthakis, J. (1969) : On a relation between the indices of solar activity in  
the photosphere and the corona. Solar Phys., 10 : 168.
- Xanthakis, J. and C. Poulakos (1978) : A forecast of solar activity for the  
21st solar cycle. Solar Phys., 56 : 467.

# The Variability of the Solar Ultraviolet Spectral Irradiances

J. Gethyn Timothy  
Laboratory for Atmospheric and Space Physics,  
University of Colorado,  
Boulder, Colorado 80309

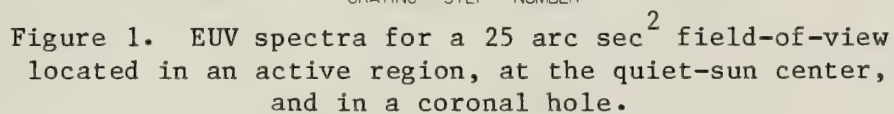
## Abstract

Available data on the solar spectral irradiances at wavelengths below 3100 Å are extremely limited and there are major uncertainties in many of the measurements. In particular, there is major disagreement on the magnitude of the variability of the spectral irradiances over the solar cycle. This paper briefly reviews our current understanding of the effects of different solar features on the ultraviolet spectral irradiances over both the 28-day solar rotation period and over the solar cycle. It is proposed that any attempt to predict the magnitudes of the ultraviolet spectral irradiances must take into account a long-term variability of emission from quiet regions of the solar disk over the solar cycle. The need for direct long-term monitoring of the ultraviolet spectral irradiances is emphasized.

## 1. Solar ultraviolet spectrum

The solar ultraviolet spectrum changes dramatically in character between 3000 and 1200 Å. Near 3000 Å the emission is primarily a strong continuum originating from the photosphere with many strong absorption lines superimposed. At about 2080 Å a sharp decrease in the continuum flux is observed, caused by the bound-free absorption edge of Al I. Other absorption edges of Mg, Fe, Ca, and Si also occur in this wavelength range. At wavelengths above about 1680 Å the spectral lines occur in absorption, and the lines and continua show limb darkening. At wavelengths between 1680 and 1520 Å, where the brightness temperature of the continuum passes through a minimum, the spectral lines are shallow and little or no limb darkening is observed. Below 1520 Å emission lines from the chromosphere begin to be observed and both lines and continua show limb brightening.

The solar extreme ultraviolet (EUV) and soft x-ray spectrum in the wavelength range from 40 to 1200 Å is dominated by strong emission lines arising from neutral atoms and ionized species in the chromosphere, transition-region and corona. Emission lines from ionization stages as high as Fe XVIII (temperature of formation about  $8 \times 10^6$  K) have been observed in non-flare spectra (c.f. Fawcett 1974), while lines of lithium-like Fe XXIV (indicative of non-thermal process of extremely high energy) have been recorded during flares (Purcell and Widing, 1972). The hydrogen Lyman continuum and the He I and II continua are also observed in emission. Examples of spectra from 300 to 1350 Å recorded with the Harvard College Observatory EUV spectroheliometer on the Skylab Apollo Telescope Mount (ATM) (Reeves, Huber, and Timothy 1977) are shown in Fig. 1. Differences in the shape of the spectrum for radiation emitted from active regions, average quiet regions on the solar disk, and coronal holes can be seen clearly in these data.



281



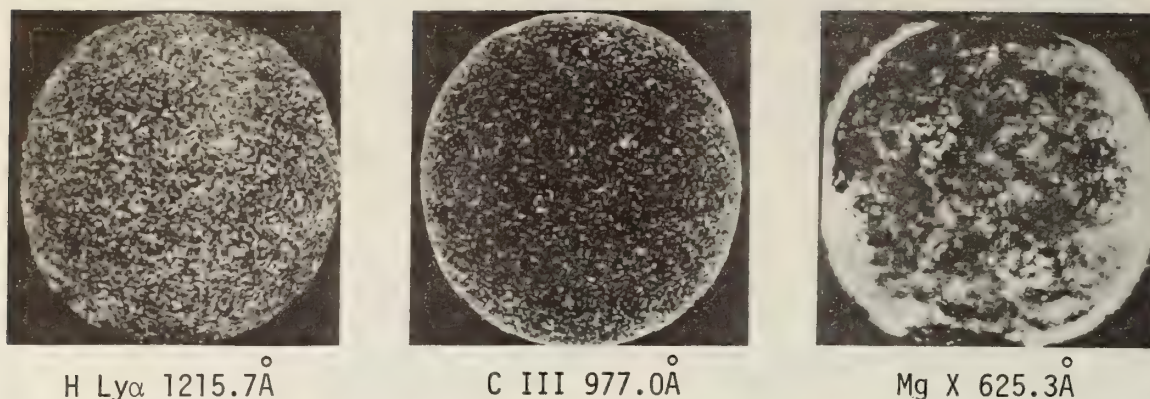


Figure 2. EUV images of the solar disk at wavelengths corresponding to chromospheric (H Ly  $\alpha$  1215.7 Å), transition-region (C III, 977.0 Å), and coronal (Mg X, 625.3 Å) emission lines.

## 2. Solar ultraviolet spectral irradiances

The current state of knowledge of the solar ultraviolet and soft x-ray spectral irradiances has recently been reviewed by Heath and Thekaekara (1200 to 3000 Å), Vidal-Madjar (H Ly  $\alpha$  1216 Å), Timothy (300 to 1200 Å), Manson (10 to 300 Å), and Walker (10 to 300 Å) in the volume "The Solar Output and its Variation" edited by O. R. White (1977). An intercomparison and compilation of solar irradiance data relevant for aeronomy is also being prepared by Delaboudiniere *et al.*, (1978). We can summarize the key conclusions of these reviewers as follows:

1. The best available irradiance data have photometric accuracies ranging from about  $\pm 10\%$  near 3000 Å to more than a factor of two below 200 Å. Typical accuracies at 300 Å are about  $\pm 30\%$  and at 1200 Å are about  $\pm 20\%$ .
2. There is almost total disagreement on the magnitude of the variability of the irradiances over the solar cycle. Heath (1978) presents data to show that the solar irradiance at 1750 Å is a factor of 2.5 greater at solar maximum than at solar minimum, with a variability at 3000 Å of 18%. However, Simon (1978) states that a variability of a factor of 2 at 2000 Å seems to be too high. Vidal-Madjar (1978) finds a variability in the H Ly  $\alpha$  1216 Å irradiance of a factor of two over solar cycle 20. The data of Timothy *et al.*, (1972) also tend to support the contention of Vidal-Madjar of a variability of about a factor of two at extreme-ultraviolet wavelengths. Conversely, Heroux and Hinteregger (1978) find for the wavelength range 300 to 1220 Å during solar cycle 20 that "...it is possible that the EUV fluxes varied by more than 10% indicated by their (Heroux and Higgins) measurements, but it appears unlikely that the extreme-ultraviolet flux variations exceeded 30%."

3. Ground-based indices of solar activity, such as the 2800 MHz radio-noise index, are highly unreliable indicators of the magnitude of the solar ultraviolet and soft x-ray irradiances (c.f. Timothy and Timothy, 1970), and must be replaced by direct measurements on a routine basis.

### 3. Effects of specific solar features on the ultraviolet spectral irradiances

Our current knowledge of the effects of specific solar features on the ultraviolet spectral irradiances can be summarized as follows:

#### Flares:

Dynamic events such as flares can cause enhancements of a factor of two or greater in chromospheric and transition-region line irradiances and of more than an order of magnitude in coronal emission-line irradiances, in a time scale of a few minutes. Although, flares occur with greater frequency during the maximum of the solar cycle, the effects are short-term and are not directly relevant to the problem of predicting long-term changes in the magnitude of the ultraviolet spectral irradiances.

#### Coronal Holes:

These are the regions of very low intensity at coronal wavelengths which have lifetimes of several solar rotations. Examples of coronal holes can be seen near the poles in the MgX (625.3 Å) image in Fig. 2. These regions have a negligible effect on the irradiances of chromospheric and transition-region lines but contribute to the variability of coronal lines.

#### Bright Points:

Bright point sources of EUV and x-ray radiation have been observed to be uniformly distributed across the solar disk (see Fig. 2). The integrated effect of these bright points is not significant in terms of the total irradiance at ultraviolet wavelengths even for coronal emission lines. However, the possibility that a variation in the number of bright points in the chromospheric network may cause a long-term change in the intensity of the emission from the quiet-sun during the solar cycle need investigation.

#### Active Regions:

The emission at all ultraviolet wavelengths is enhanced in the dense plasma of active regions. The observed ratios of the average active-region and quiet-sun intensities for specific extreme-ultraviolet emission lines are shown in Fig. 3 (Noyes, Withbroe and Kirshner, 1970).

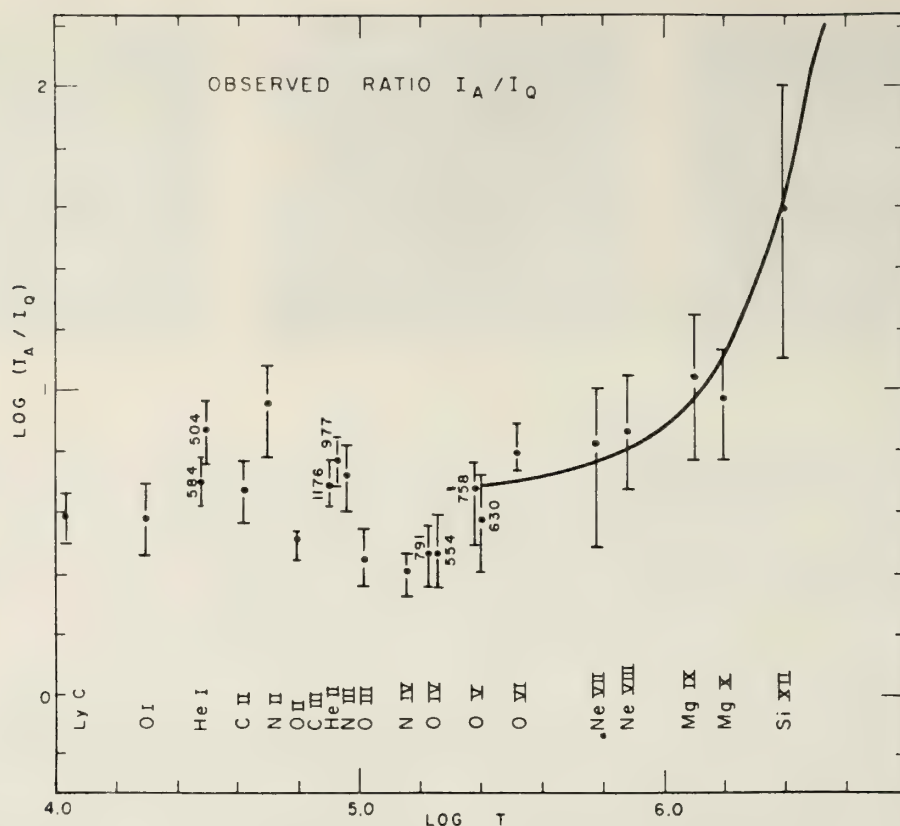


Figure 3. Mean enhancement of EUV emission from active regions ( $I_A/I_Q$ ) where  $I_A$  = average intensity per unit area in active region and  $I_Q$  = average intensity per unit area in quiet regions of the solar disk (Noyes, Withbroe and Kirshner, 1970).

The effect of the active regions is to cause variations of about 1% at 3000 Å and 6% at 1750 Å during the solar rotation period. A variation of about 30% is observed in chromospheric and transition-region lines at extreme-ultraviolet wavelengths, while variations of up to a factor of three have been observed in coronal emission lines. No systematic change in the magnitude of the variability during a solar rotation period has so far been observed from the maximum to minimum phases of the solar cycle. These observed variations are apparently inadequate to account for the magnitudes of the variabilities at ultraviolet wavelengths over the solar cycle that have been reported by some investigators.

#### Quiet Sun:

The quiet sun is dominated by the chromospheric network which has a well-defined structure when observed at ultraviolet wavelengths characteristic of temperatures in the range  $10^4$  to  $10^6$  K. At coronal altitudes ( $T > 1.5 \times 10^6$  K) the network is no longer visible and the majority of the radiation originates in active regions and large scale magnetic loop structures.

Analyses of the extreme-ultraviolet images of the quiet sun recorded with the Skylab spectroheliometer, which are now being completed, show significant change in the relative brightness of the chromospheric network



and the network cells as a function of time. As shown in Fig. 4., the ratio of the average quiet sun intensity to the cell intensity declined during the period of declining solar activity in 1973. A reduction in the absolute value of the quiet sun intensity, of a magnitude greater than the experimental error, was also observed during this period. It can be seen in Fig. 4 that the greatest contrast, and the greatest change of contrast, is observed for lines characteristic of transition region temperatures.

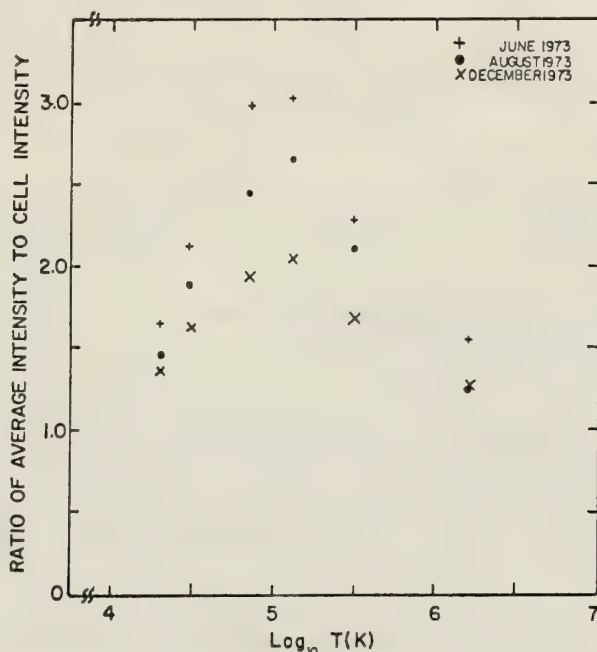


Figure 4. Variation of the ratios of the average quiet sun intensities to the network cell intensities as a function of the temperature-of-formation of the extreme-ultraviolet emission line.

Hinteregger (1977) reports changes in the irradiances of chromospheric emission lines during 1975 for periods when the sunspot number was zero, with an apparent "EUV solar minimum" in April 1975. Subsequently, between 1976 and 1979, the irradiances appear to have increased dramatically (Hinteregger, private communication). The combination of these different data sets strongly suggests a major variability in the intensity of the ultraviolet emission from the quiet sun over the solar cycle.

#### 4. Prediction of solar ultraviolet irradiances

There is at present insufficient data to permit any attempt at predicting the magnitudes of the solar ultraviolet spectral irradiances during the solar cycle. Clearly there is a critical need for accurate measurements of the spectral irradiances at wavelengths below 3000 Å in order, first, to determine the magnitude of the variability as a function of wavelength and, second, to attempt to relate the "ultraviolet solar cycle" with conventional ground-based indices of the level of solar activity. However, based on the analyses of the Skylab data, I propose that a prediction model for the

variability of the ultraviolet spectral irradiances will need to consist of two components. First, an 11-year periodicity associated with the variability of the emission from the quiet sun and, second, a 28-day periodicity associated with the appearance and disappearance of active regions at "active longitudes" on the solar disk.

#### References

- Delaboudiniere, J. P., R. F. Donnelly, H. E. Hinteregger, G. Schmidtke and P. C. Simon, COSPAR Technique Manual Series 7, (1978).
- Fawcett, B. C., Adv. Atom. Molec. Phys. 10, 224 (1974).
- Heath, D. F. and M. P. Thekaekara, in "The Solar Output and its Variation," ed. O. R. White, Colorado Associated University Press, 1977, p. 193.
- Heroux, L. and H. E. Hinteregger, Jour. Geophys. Res. 83, 5305 (1978).
- Hinteregger, H. E., Geophys. Res. Lett. 4, 231 (1977).
- Manson, H. E., "The Solar Output and its Variation," ed. O. R. White, Colorado Associated University Press, 1977, p. 261.
- Noyes, R. W., G. L. Withbroe, and R. P. Kirshner, Solar Physics 11, 388 (1970).
- Purcell, J. E. and K. E. Widing, Astrophys. J. 176, 239 (1972).
- Reeves, E. M., M. C. E. Huber, and J. G. Timothy, Applied Optic 16, 837 (1977).
- Simon, P. C., "The Solar Output and its Variation," ed. O. R. White, Colorado Associated University Press, 1977, p. 212.
- Timothy, A. F. and J. G. Timothy, Jour. Geophys. Res. 75, 6950 (1970).
- Timothy, A. F., J. G. Timothy, A. P. Willmore and J. H. Wager, J. Atoms. Terr. Phys. 34 969 (1972).
- Vidal-Madjar, A. in "The Solar Output and its Variation," ed. O. R. White, Colorado Associated University Press, 1977, p. 213.
- Walker, A. B. C., in "The Solar Output and its Variation," ed. O. R. White, Colorado Associated University Press, 1977, p. 279.

## B. SHORT-TERM SOLAR ACTIVITY PREDICTIONS

A WORKING GROUP REPORT prepared by P. Simon, Chairman, and J. B. Smith, Jr., Co-chairman, and Working Group Members: Y. Ding, W. Flowers, Q. Guo, K. L. Harvey, R. Hedeman, S. F. Martin, S. McKenna Lawlor, V. Lin, D. Niedig, V. N. Obridko, H. Dodson Prince, D. Rust, D. Speich, A. Starr, and N. N. Stepanyan.

### 1. HISTORY

The first broadcast of geophysical data started on December 1, 1929, at the Eiffel Tower Station, at Paris (France) "in order to study the geomagnetic storms and their related phenomena". In a few years, several other countries began to establish such broadcasting programs. In the fifties, all the relevant Warning Centers began to exchange data and to cooperate inside the URSIGRAM network established jointly by wireless companies and sponsored by URSI (Union Radio Scientifique Internationale). The main effort was to agree internationally on the codes used in data exchange and to improve the forecasting techniques which were entirely devoted to the short term forecast of ionospheric propagation and disturbance. Just before the sixties, with the International Geophysical Years (1957-1958), they evolved toward a solar forecast devoted to scientific programs. With the coming of space research, scientific equipment on rockets, balloons and satellites required forecasts and warnings in order to perform their experiments successfully. Simultaneously the forecast itself evolved. Starting during IQSY (1964) solar astronomers issued global forecasts of solar activity in two categories: quiet and active. During the Proton Flare Project (1966) proton events were successfully forecast at Meudon, where daily forecasting was initiated for the activity of each spotted group on the disk according to four categories: QUIET, ERUPTIVE, ACTIVE and PROTON. At about this time, the Space Environment Services Center, utilizing real-time reports of solar activity from newly established flare patrol observatories operated by the National Oceanic and Atmospheric Administration (NOAA), with NASA support, and the U. S. Air Force, in addition to the cooperative scientific observatories with flare patrols, began issuing daily predictions and discussions of solar activity and geophysical responses. In the field of geomagnetic disturbances, forecasts were improved by the use of radio burst data and, more recently, coronal hole reports. At the present time, the forecasting of solar activity at several centers around the world is being done on a routine basis incorporating a wide range of both ground-based, as well as satellite-based, solar observations. Forecasts are generally issued as real time reports on expected events and prompt information is provided about unexpected events in progress.

Accurate forecasting requires not only the input of up-to-date solar data, but also a fundamental knowledge and understanding of solar activity processes. The purpose of this report is to present the current status of solar activity forecasting. Our discussion summarizes the forecast centers



and the specific applications for which the solar activity forecasts are used, the scientific basis used in making forecasts and specifically how, as well as how well, these forecasts are made using systematically monitored solar data. Recommendations, which this working group suggests will improve the forecasting of solar activity, are also presented.

## 2. FORECASTING CENTERS AND THEIR CUSTOMERS

### 2.1 IUWDS Regional Warning Centers

Presently forecast centers are operating as Regional Warning Centers within the framework of the International URSIGRAM and World Days Service (IUWDS); each has its own National program but they closely cooperate in international programs.

The key components of a forecasting center are (a) an observation program in order to have their own set of data, (b) forecasters, and (c) communications for receiving and distributing data and forecasts. Most of the forecasting centers are operated seven days a week during business hours; a few operate five or six days a week. Some operate as much as sixteen hours daily (e.g. Boulder).

The IUWDS Regional Warning Centers currently operating are the following: (1) Forschungs Institut der DPB, Darmstadt, Germany FR, (2) Institute of Applied Geophysics, Moscow, USSR, (3) Ionospheric Prediction Service, Sydney, Australia, (4) Meudon Observatory, France, (5) Radio Research Laboratory, Tokyo, Japan, (6) Space Environment Services Center, NOAA, Boulder, Colorado, USA. The level of activity of Forecasting Centers (FC's) roughly follows the solar cycle. At sunspot minimum, very few groups are really concerned with solar forecasts; but as soon as the sunspot number significantly increases, the FCs begin receiving a large number of requests. Of course, many requests emanate from people whose involvement (interest) fades out after a short time.

### 2.2 Categories of Customers

We can identify three categories of customers.

- (1) Those actually using a forecast in order to begin or to continue a program of operation, for example, the following:
  - a) Integrated-observation programs devoted to the study of transitory events (flares, type III metric burst, sudden commencement magnetic storms) and of evolving processes (flare build-up, evolution of solar active centers, etc.). We have in mind the 1966 and 1969 Proton Flare Projects, the 1972 Campaign for Integrated Observation of Solar Flares, the coordinated program of the Incoherent Scatter Facilities, the Skylab/ATM Solar program, etc. Both the necessary delay involved in contacting all the relevant stations and the need for recording and detecting the onset of events require a program of observation based on forecasts issued one or several days in advance.

- b) Automatic experiments operated aboard a gondola of stratospheric balloons. Most of the time, this equipment cannot be recovered for a second flight and consequently the experimenter can only use a limited amount of equipment during his observations program. The duration of the flight is limited - ten to twenty hours to several days. The delay between the decision for flight and the experiment operation at ceiling altitude is several hours. In addition, scientists may be interested in the beginning of an event (magnetic storm, solar proton arrival, etc.) or in having balloons scattered in a large geographical area at the time of occurrence of an event. Therefore, some stratospheric balloon experiments require solar-geophysical forecasts for flight decisions.
  - c) Satellites with instruments that either have several modes of operation or are operated in a series of joint observing sequences, or have a solar telescope with a narrow field of view, e.g. 1 arc degree, or 10 arc seconds. These satellite experiments are computer operated and their program is usually selected one or several days in advance. During the operation, the experiments can only change either the telescope pointing or a few minor details of the programs. These satellite experiments must be operated according to a forecast of solar activity.
  - d) Ground based programs which can only be carried out during specific geophysical or solar conditions. We have in mind the geological study of deep layers of the Earth by seismologic process and telluric currents. This work is carried out on the ground or aboard low speed aircraft towing the relevant instrument, and is used for oil and water supply research. Having a forecast saves time and expense.
- (2) Those using a forecast as a warning: experimenters waiting for additional data in order to start their program of operation, for example, the following:
- a) Rocket experimenters. They cannot keep their rockets on standby for long intervals of time, but rather only when they have a positive forecast. However, the firing decision will be made according to data acquired at the rocket range (magnetogram, riometer, real-time flare report or satellite data, etc.) or to direct observation of an event in progress (aurora, etc.).
  - b) Satellite experimenters. The forecast is used to verify that the appropriate programs can be performed at the relevant time. For example, satellites with instruments which cannot be run properly during a magnetic storm.)
  - c) Staffs operating transmitters using ionospheric propagation: they must distinguish between a "fade out" and a sudden breakdown of their transmitter.
  - d) Engineers controlling the electric power distribution at high geomagnetic latitude are interested in geomagnetic storm forecasts:



they do not follow the same scheme for restoring the distribution in the case of a geomagnetic-storm generated overload as in the case of an overload caused by an accident.

(3) Those using a solar forecast in order to prepare another forecast, for example, the following:

- a) Ionospheric propagation. The forecasters require a successful forecast of the geomagnetic storm and of the time of its beginning in order to forecast ionospheric propagation disturbances.
- b) Satellite drag. The space agencies use solar and geomagnetic forecasts in order to predict atmospheric density and, hence, the trajectory of the low orbiting satellites and their re-entry in the Earth's atmosphere.
- c) Balloon experimenters. Some balloon experimenters are interested in substorm observations. Balloons are launched after the beginning of the geomagnetic storms, using local magnetic data in order to predict the substorm occurring late in the night.

One can easily understand that all these users do not have the same requirements. Several of them need a "yes or no" forecast, while others would appreciate supplementary pieces of information. Putting this kind of question to a forecaster implies a personal confidence in his judgment; the forecaster's reply requires a careful survey of all the elements entering the problem.

### 3. SOLAR ACTIVITY

Observed structures in the solar atmosphere, for example, filaments, coronal holes, and active regions, result from the structure and evolution of the magnetic fields on the Sun. The source of these fields is sub-photospheric. The fields rise through the photosphere and above producing active regions. The frequency of region formation specifies the course of the solar cycle and the subsequent development, interaction and dissipation of the active-region fields result in the many solar structures and conditions known to ultimately affect the Earth. Understanding these processes will allow us to effectively predict solar activity over short and long term periods.

We summarize our current understanding of solar activity used as the basis for predicting solar activity in three sections (1) active centers (AC), including what is known about active regions and their development over a period of a few days to several months; (2) flares and associated effects, including the known relations between the observable active-region parameters and both the time and location of solar flare occurrence; and (3) the solar origins of geomagnetic activity, including a summary of the observed conditions and structures on the Sun which give rise to geomagnetic activity.



### 3.1 Active Regions

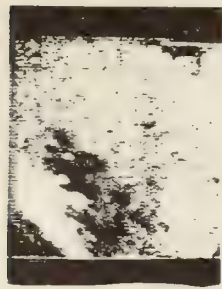
In the monitoring and prediction of most transient solar phenomena, the active region<sup>AR</sup> condition and evolution is of primary concern. As used in this report, the AR is defined as the identifiable magnetic entity within which most of the field lines are footed. An AC is defined as the complex of more than one AR which are close spatially and apparently strongly interactive magnetically. The following are significant examples: (1) Essentially without exception, flares occur in centers of activity, where magnetic field strengths are enhanced, chromospheric plage is visible and, usually, sunspots are visible. It is also true that major flares (with the well-known spotless region exceptions) show a strong preference for regions with complex magnetic structures, bright and compact plage, and large sunspot groups. It is therefore evident that the prediction of major flares, particularly for periods  $\geq 1$  day, depends largely upon correctly predicting the evolution of the active region(s). The magnetic structure and evolution of nearby active regions during their disk transit are shown in Figure 3.1. (2) The correct prediction of the full-sun X-ray, EUV, or 10 cm flux, which are important in atmospheric modeling, communications, etc., is dependent on the correct forecasting of the contribution to that flux by the individual AR's. (3) The correct prediction of the formation evolution of Coronal Holes is probably dependent on the correct prediction of active region evolution. (4) Generally speaking, any specific evolution which is relevant to the solar-interplanetary-terrestrial relationship, such as coronal transient occurrence, the interplanetary magnetic field structure and evolution is related to the evolution of active regions.

#### 3.1.1 Active Region Evolution

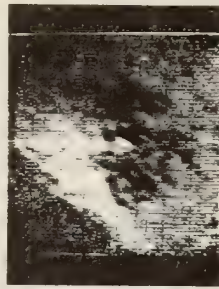
Svestka (1976a) recently described active-region evolution in detail. As soon as new magnetic flux emerges from subphotospheric layers of the sun into the solar atmosphere, the solar plasma is heated and an "X-ray bright point" is formed. Some of the bright points that live longer continue to grow and strong magnetic flux continues to emerge nearby. The original bright point transforms into a long-lived and much more extensive area of strong magnetic field in the atmosphere called a solar active region. The brightening in soft X-ray and in microwave emission indicates a temperature and density increase in that part of the corona where the magnetic field is strengthened. The chromospheric plages resembles closely the distribution of the photospheric magnetic field, while the coronal manifestation is more diffuse. An active region grows through subsequent emergence of new flux tubes, and the basic structure of an active region is a system of loops. In a well developed region, however, the loops merge and it is difficult to distinguish them in the homogeneously bright, hot X-ray cores.

The growth phase is the most active phase of the region development. Individual loops, or systems of loops in the active regions, are suddenly heated to temperatures of tens of millions of degrees, and in some cases elementary particles are accelerated in such loops to hundreds of keV and to even higher energies. Accelerated electrons cause hard X-ray bursts and radio emission. Energy is conveyed to the chromosphere, which is thus excited at the foot-points of these loops. The major flares occur only after the sunspots are formed, close to the peak of development of the active region.

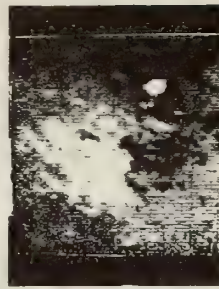
# MAGNETIC EVOLUTION AR 1374/1381



30 OCT. 1978



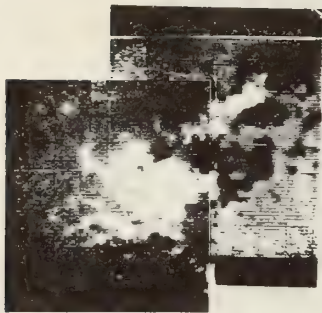
31 OCT. 1978



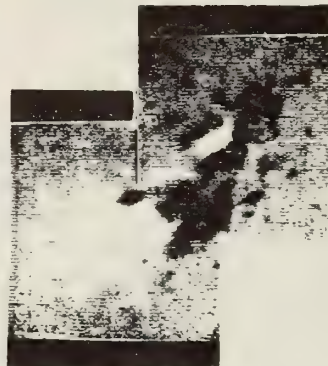
1 NOV. 1978



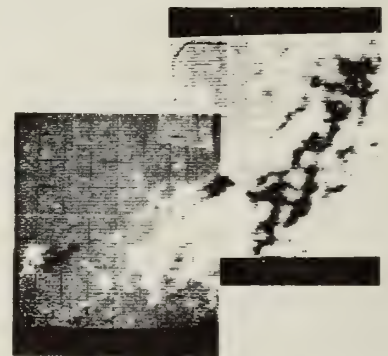
2 NOV. 1978



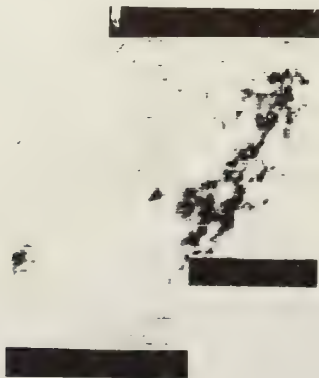
3 NOV. 1978



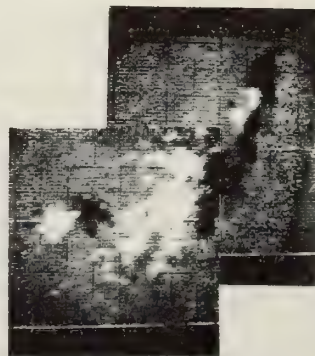
4 NOV. 1978



5 NOV. 1978



6 NOV. 1978



8 NOV. 1978



10 NOV. 1978

Figure 1. The longitudinal fields of an active region complex is shown for its disk transit. Note the birth of an interacting active region on 1 November and the appearance of a small trailing bi-polar region on 5 November. (North is upper left, east lower left. Negative fields are dark and positive fields are light). The magnetograms are courtesy the Marshall Space Flight Center.



During the decay phase of active regions, the gas inside the loops becomes condensed, it cools, and then it appears in the form of a dark quiescent filament as seen in the optical  $H\alpha$  line. This process starts when the active region is quite young, but the filament becomes most pronounced after the peak of the active region development, when the active region starts to decay. First the sunspots disappear, then the coronal condensation becomes less dense, the bright compact core dissolves, the radio emission decreases, and we again begin to see the loop structure more clearly, with the loops connecting the opposite magnetic polarities. The configuration of hot loops embedding a cool filament, born earlier and now slowly decaying, can be seen as a remnant of the active region sometimes for many months. All structures seen in the corona may be remnants of disintegrated activity. Finally in the late phase of development, the filament disappearance still produces outstanding enhancements in the coronal X-rays and, occasionally, spectacular chromospheric flares.

The X-ray photographs of the sun have also revealed that the extension of active regions in the corona is much larger than the underlying sunspot groups and plages may indicate. During the Skylab mission about 100 loops were identified as interconnecting separate active regions. Thus the length of the interconnecting loops indicates the actual extent of relatively strong fields associated with active regions. More than half of the active regions extend farther than  $20^\circ$  from their cores; their size is about  $40^\circ$  in diameter and sometimes even larger.

The individual active regions we see on the sun are often interconnected, up to distances of some  $30^\circ$ , perhaps not only in the corona where we may observe the connections in X-rays, possibly also below the photosphere. Activity of several such nearby regions then can have a common ultimate source. The regions may have many features in common, and we should not treat them individually, but rather in the broader context of a complex of activity.

Svestka's description shows well the complexity of an active region but it should be completed by including coronal streamers, the structures shown by the metric and decametric observations, etc.

### 3.1.2 Prediction of Active Region Evolution

The forecaster bases his prediction of the evolution of an active region on a set of data including magnetic field measurements and observations at white light (sunspots),  $H\alpha$ , and X-ray wavelengths. Currently only full disk X-ray measurements are routinely available but spatially resolved soft X-ray data are very desirable. Forecasters also use spatially resolved radioheliograph observations at wavelengths ranging from microwave to the metric and decametric wavelengths. Many of these observations including the satellite observed X-rays are free from interference by the weather conditions at the forecast center.

The forecaster depends upon the above data to predict the evolution of the active regions and the resultant solar activity (and geographical responses). The forecast interval can be divided roughly into two categories: (a) Predictions a month or more in advance, which are <sup>made</sup> by considering the large



scale evolution related to the persistence of active and quiet areas on the sun, including active or "preferred" longitudes and the phase of the overall solar cycle (The evolution of large scale patterns in the solar atmosphere also appear to be related to active region development (McIntosh, 1979) and this is currently being used in the prediction of active region development.) and (b) Predictions for one or a few days in advance. The more detailed short term forecast is related to the evolution of the individual-active regions and complexes of active regions.

### 3.1.3 Active Longitudes and "Long-Term" Prediction

For intervals of several months or longer, high levels of flaring may repeatedly occur at the same solar longitudes. It has not been established with certainty whether the nature of these active longitudes is due solely to the long lifetime of a few large active centers or complexes of active centers or whether it is also due in part to the preferred occurrence or growth of long lived active centers or complexes of active centers at selected longitudes. Even though the positions of the birth of active centers in latitude is highly systematic (as revealed by the "butterfly" diagrams), it has not been statistically shown that the positions of birth of active centers in Carrington longitude is anything other than random. However, the prevailing consensus based on morphological studies is that active centers have a definite trend to emerge as "families" or "streams" very clustered at preferred longitudes.

In addition to the possibly significant clustering of major active centers at preferred longitudes there is also the possibility that those new active centers, which develop within existing active centers, may interact in such a way as to prolong the lifetime of either the old active center, the new active center, or both. Soft X-ray images obtained by the AS&E/MSFC experiment and the NRL UV experiment on board Skylab confirmed the existence of loops or arches interconnecting nearly all adjacent active centers. Hence it is suspected, although not proven, that the magnetic interconnections between groups of adjacent or intermingled centers may influence the development and prolong the lifetime of such a complex of active centers.

The repeated occurrence of flares within complex active centers is of importance to the flare forecaster irrespective of whether the active center is, or is not, located near the active longitudes. A question relevant to the forecaster is "How reliably can the knowledge of the repeated occurrence of flares at active longitudes be used in forecasts?" The reliability of such forecasts depends upon both the width of the active longitude "window" and the length of time into the future for which the forecast is being made. For example, during the maximum of solar cycle 19, active longitudes were identified at Meudon for 75 rotations beginning in 1956. During this interval, there were five to six times more flares in the active longitude zones as in the remaining longitudes. For the most active longitude ( $20^{\circ}$  wide) the number of Polar Cap Absorption (PCA) and/or Type IV events was 13 times as great as in the non-active longitudes. During the Skylab Mission, a much less active period, almost all solar flares occurred in narrow longitude belts, which were on one hemisphere of the sun. In real time, of course, the forecaster does not have at his disposal prior knowledge of active longitudes (as above). However, if a forecaster always predicted that a PCA and/or Type IV would occur in the next rotation after the first occurrence of a PCA or Type IV, he would be correct in one out of every four predictions.

### 3.1.4 Short-Term Active Region Evolution and Prediction

The short-term evolution of active regions and active centers is very familiar to solar observers and forecasters. During the workshop, forecaster members of the working group individually analyzed the same data set and made predictions for the following day. The predictions were strikingly similar and correctly reflected the subsequent activity. However, only a few systematic studies have been carried out. Some studies have related solar activity to the sunspot classification. The more detailed McIntosh modified Zurich sunspot classification was correlated with flare production by Kildahl, and is published in a later volume of these proceedings. However, magnetic field observations of the plage of active regions have shown that the magnetic (sunspot) structure, as described by the Mt. Wilson classification, can ignore the actual complexity of many active centers, which play an important role in the flaring activity of the active regions. Martres (1968) studied the mutual influence between active regions located nearby. Martres, et al. (1973) also studied the relationship between the photospheric velocity field and the evolution of active centers. The close relationship between the velocity and magnetic field (including the coronal structure) must be further studied to understand the physical processes involved in the appearance and development of active regions. The relationship between the magnetic structure of active regions and their level of emission in X-rays, UV, and at radio wavelengths also requires further study. More accurate forecasts of UV emission, even for a few days in advance, would be valuable for many purposes.

### 3.2 Flares and Associated Events

Observational studies have demonstrated the close relation between flare activity and the magnetic field of active regions. Theoretical work further suggests that the storage of sufficient energy for flares can only occur in the magnetic field of an active region and most likely in the fields in the low corona (Rust, 1976a,b). How this energy is accumulated depends on the course of evolution of the active region. The introduction of new flux into a region and the motions and rotations of sunspots play an important role in establishing the state of higher magnetic energy found in flare-producing regions. In some instances, photospheric motions may even provide sufficient energy for flares (Sturrock, 1972; Heyvaerts, 1974; Priest and Raadu, 1975).

The detection and monitoring of the energy build-up process in the magnetic fields in the low corona and of those conditions leading to the release of the stored energy (the flare) would allow the most optimum and effective anticipation of flare occurrences. However, this has not as yet proved possible. Therefore, to predict when and where flares may occur (and in some cases the energetics of a flare) we currently must rely on recognized relations of flares and observed parameters of active regions. In the following discussion we present these relations in two general sections: (1) the observed identification of flare sites based on the evolution and configuration of active regions, (2) the identification of changes in the region's structure prior to and during flare occurrence which may specify when flares will occur. Flare energetics are briefly considered. Because of several excellent reviews of solar flare research, we have not attempted to present a comprehensive report of the literature. We consider the overall view of what we know



of where and when flares will occur referring the reader to Svestka (1976), Rust (1976a,b), Martin (1979), Smith (1972) and the Skylab Flare Workshop report (1979) for further details. References cited herein are for the most part those not discussed in these reviews.

### 3.2.1 Magnetic and Velocity Field Configurations Identifying Flare Sites

The association of increasing flare activity with increasing complexity of the magnetic field of active regions was first established by Giovanilli (1939) and has been verified in many subsequent studies. Complex bipolar groups and regions with multiple bipolar components produce three times as many flares as do simple bipolar regions. The most magnetically complex regions also produce the most important flares, with the highest and often the most energetic-flare production occurring in reversed-polarity regions. Significant proton flares are generally associated with  $\delta$ -configuration sunspot regions (Tanaka, 1979; Ding et al., 1979) and often in a region with an inverted magnetic axis (Tanaka, 1979; Zirin, 1970). It is equally noteworthy, however, that a number of proton flares (though generally weaker events) have occurred in extensive simple bipolar regions with few or no spots (Dodson and Hedeman, 1970, 1979). In many cases, these isolated proton events occur in regions that are the simplified remnants of previously complex regions and are usually related to the disruption of a large prominence.

The interaction of two or more close regions also appears to be an important factor in flare occurrence. The mechanism for this seems to be the magnetic interconnection of active regions. Skylab X-ray-image observations and Culgoora and Nancay interferometric radio studies have demonstrated that the magnetic fields of optically distant regions can be connected.

The association of flares to specific patterns in the magnetic field and changes in those patterns has been studied extensively over the last few years. These investigations have used (a) direct observations of the photospheric magnetic field, (b) the inferred chromospheric magnetic field, i.e. inferred from the optical structure, and/or the coronal magnetic field -- inferred from X-ray imaging data. While no one observed parameter or set of them has yielded necessary and sufficient observational conditions for the occurrence of a flare, a general picture is forming which allows us to recognize with some statistical certainty potential flare sites and also to aid in developing the theoretical models necessary to understand the flare process.

Flares almost invariably occur along a magnetic inversion line, i.e., where the vertical magnetic field is zero. Magnetic inversion lines most frequently result from the four active region evolutionary circumstances illustrated in Fig. 3.2. Every active region has a basic large scale bipolar structure and hence a primary magnetic inversion line when one views the vertical component of the magnetic field. Flares may occur around this primary magnetic inversion line as illustrated in (A) in the schematic diagram of typical magnetic region configurations in Figure 3.2. A second location where some flares occur is around magnetic inversion lines which result when two or more active regions develop very close to each other, (B). A third magnetic inversion site for flares forms whenever a new active region develops within a pre-existing active region (C). A fourth type of evolutionary circumstance which results in the development of magnetic inversion lines is the formation of satellite poles around the periphery of large sunspots (D).



For a flare to occur other conditions must be present. Observable conditions frequently associated with flares are (a) the apparent convergence and interaction of adjacent active regions and of network flux, (b) satellite sunspots, (c) the emergence of new flux in an existing region, (d) a shear in the magnetic and/or velocity field of a region (e) strong gradients across the magnetic inversion line.

a) Converging Magnetic Flux. The "collision" of adjacent and developing active regions has been found to result in major flare production. This process also appears to occur on a smaller scale with the apparent convergence of network flux of opposite polarities or of part of an active region with old region flux of opposite polarity (Harvey and Martin, 1979) and when ephemeral active regions expand into neighboring network fields (Marsh, 1978).

b) Satellite Sunspots. Satellite sunspots or poles are observed as small features of opposite polarity adjacent to a sunspot. They may or may not have a pore. Small flares are often observed with growing or decaying satellite spots and frequently are accompanied by surges. Generally no flares occur if the satellite sunspot is static.

c) Emerging Flux Regions and Evolving Magnetic Features. The emergence of new magnetic flux within an active region appears to be a critical factor in the production of flares. The emerging flux may be identified as an area of one polarity embedded in a larger field of opposite polarity or as a distinct new bipolar region. The level of flare activity is related to the position of the emerging flux in the region and to the amount of magnetic flux in the emerging regions. The likelihood of flare occurrence increases with a larger emerging flux region or if the new flux erupts as the magnetic inversion line, especially if strong gradients are present. McKenna-Lawlor (1979), for example, finds that small umbrae bordering the magnetic inversion line are especially flare related. Newly emerging flux may be directly related to the occurrence of most major flares and in fact may act as a flare trigger. Also indicative of this is the observation of rapid changes (growth and decay) occurring in spots underlying major flare sites reported by many investigators (e.g., Severny, et al., 1979). Martres et al. (1974, 1970) have found a good correlation of flare sites to "evolving magnetic features". Flares occurred at locations of two adjacent, opposite polarity magnetic structures, one of which was increasing in flux and the other decreasing. The observed changes in flux ( $10^{15} - 10^{16} \text{ Mx s}^{-1}$ ) were consistent with that found for satellite sunspots associated with flare activity.

d) Magnetic and Velocity Shears. Recent observational and theoretical investigations have increasingly indicated the importance of magnetic shears at flare sites. Their importance lies in their energy storage capacity in the fact that shears are identifiable by certain chromospheric and photospheric patterns in optical observations. Shears result from sunspot proper motions and rotations which have been found to be associated to flare occurrence. Active regions, for example, having sunspots showing large rotations are often proton-flare producers. Sunspot motions observed before the occurrence of major flares often decrease or cease after a flare.

## FLARE SITES ALONG MAGNETIC INVERSION LINES

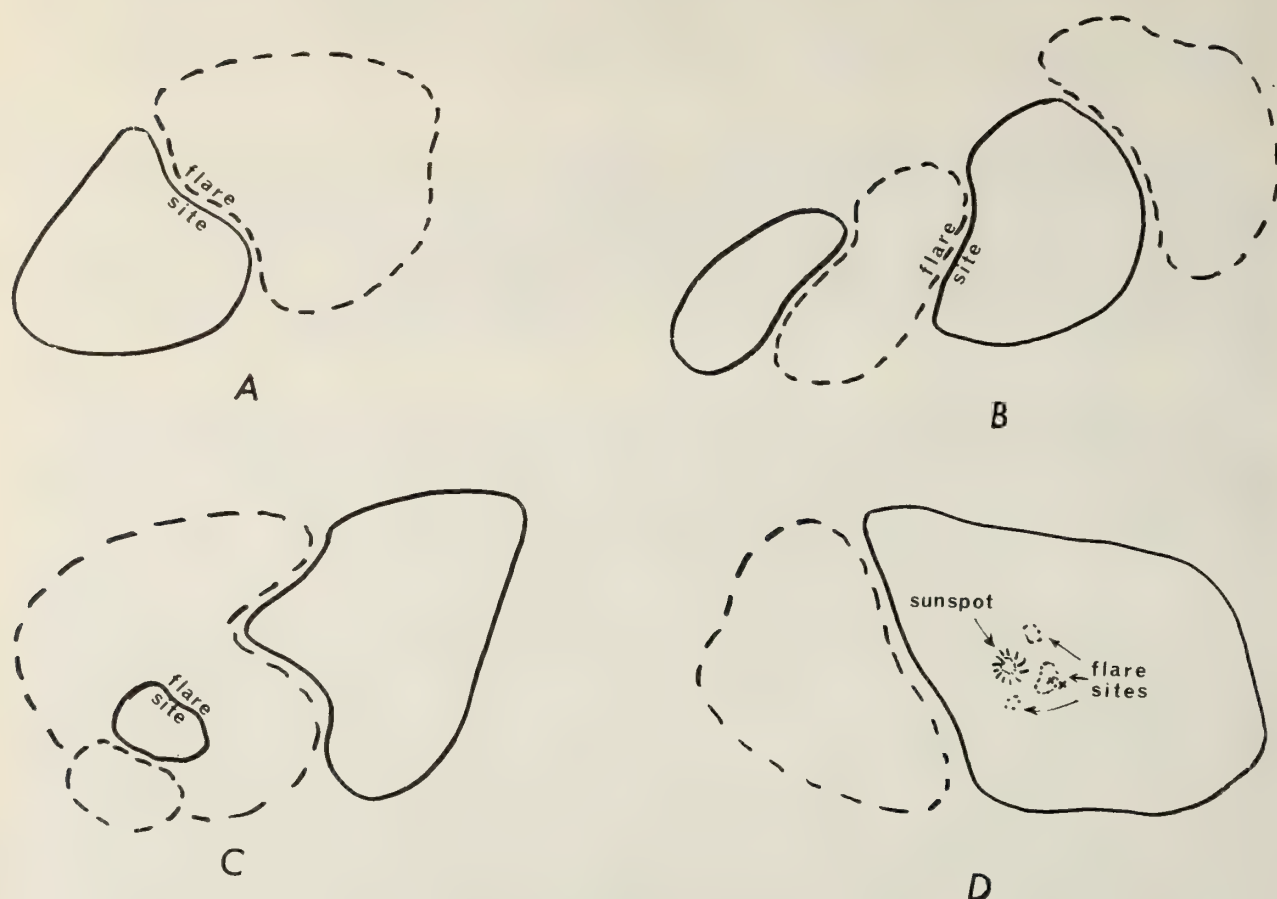


Figure 3.2

Four examples of the formation of magnetic inversion lines.

That a magnetic shear exists at flare sites has been demonstrated by several observations: (i) the direction of transverse fields at flare locations is observed to be at small angles or parallel to the magnetic inversion line, (ii) the existence of a filament along the magnetic inversion line, (iii) spiral structure observed in the penumbral filaments (Ding et al., 1979a, b; Tanaka, 1979), and (iv) the fibril orientation lying approximately along the magnetic inversion line. Changes in the fibril orientation to progressively small angles with respect to the magnetic inversion line before flares have been interpreted as a shear configuration developing (Niedig et al., 1978, Tanaka, 1979). Energy storage, calculated on the basis of the fibril orientation, has been found to be sufficient for flares (Niedig, 1978; Niedig et al., 1978).

In addition to the flare phenomenon itself, we note the role of the shearing sunspot motions in the association of the flare incidence with the magnetic neutral line ( $H_{||} = 0$ ) orientation in sunspot groups. In a normal,



flare-poor bipolar spot group, the principal neutral line is oriented roughly north-south. This, coupled with the normal sunspot motions, where the leader polarity spot moves forward and the trailer polarity moves backward or remains stationary, produces only a stretching of the field lines, without shear. On the other hand, if the neutral line is oriented east-west, or if the spot polarities are reversed, then the same direction of motion of the polarities as in a normal group will produce strong shears and lead to flares. Thus, even our means of identifying flare-productive regions in general may be partially understood on the basis of magnetic shear.

Specific patterns in the velocity field are also observed at flare sites. Though only a few studies have been made using direct velocity field observations, the velocity pattern at flare sites appears to be predominantly horizontal and can be described as a velocity shear in at least one component, (Harvey and Harvey, 1979). Similar results have been found by Martres, et al. (1977), though they interpret the observed pattern as a vortex motion. They also found that flares occurred only at locations where the velocity neutral line ( $V_{||} = 0$ ) crosses the magnetic neutral line ( $H_{||} = 0$ ), a result not completely confirmed by Harvey and Harvey (1979). For further discussion of magnetic shears and their significance in flare occurrence, see Annex I (Section ).

e) Magnetic Field Gradients. Particularly conducive to flare occurrence, certainly with the more energetic flares, is the existence of strong gradients across the magnetic inversion line. Such a condition is often observed in newly emerging regions, and can develop or strengthen as a result of emerging flux within a region, from the proper motions of sunspots, from the convergence of opposite polarity network flux from the collision of two active regions. Steepening gradients along a magnetic inversion line oriented east-west with large spots on either side is a condition invariably associated with proton flare production.

### 3.2.2 Changes in the Magnetic and Velocity Fields Associated with Solar Flares

Direct observations of changes in the magnetic field occurring near or at the times of flares have been attempted but have generally, until recent years, yielded conflicting and often negative results. The problem has been one of insufficient time and spatial resolution in the data and that the changes one might expect were comparable to the noise level of the observations. Changes in the magnetic field, however, have been observed coincident with some flares, as for example the large flares in August 1972.

That the magnetic field changes with the occurrence of a flare, as well as prior to a flare, has been inferred from the behavior of the following active region structures: (a) the expansion of arches in the corona, observed in the 5303 Å green line, on the order of 75 minutes prior to the flare onset, (b) the ascension of a filament beginning about 30 minutes, though by as much as 180 minutes, before flare onset (Eruption of the filament occurs near the beginning of the flare. About one-third of flares of importance two or greater are associated with a filament eruption. Nearly all two-ribbon flares in spotless regions are preceded by this phenomenon. In many cases, a filament will show increased activity prior to a flare but will not erupt.);



(c) a gradual increase in soft-X-ray emission prior to a flare, which is identified with the preflare filament expansion phase, (d) large proper motions of sunspots have been observed prior to proton flares while after a major flare these motions often decrease or cease, (e) changes in the fibril and penumbral structures just prior to and during flares, (f) Lang (1979) reports changes in the degree of circular polarization of small scale radio sources at the times of flares. He interprets this as indicative of a magnetic field change, and (g) calculations of the coronal magnetic field of an active region before and after the occurrence of a proton flare that suggest that the field configuration changes from closed to open as a result of the proton flare. Consistent with this is the formation or great strengthening of a small coronal hole after the proton flare of 28 April 1978 (Harvey and Sheeley, 1979). Transient coronal holes have been found in most cases to be associated with filament eruptions (coronal transients). They are indicative of a short-lived ( $< 48$  hours) change from a closed to an open field configuration resulting in the filament eruption, returning to a closed configuration within a short time (Rust, 1979).

Changes in the photospheric velocity field have also been observed at the times of some flares. Blue shifted velocities occur prior to flare start by about ten minutes through flare onset. Red shifts are observed during and after flare decay. Harvey and Harvey (1979) found that the development of a small scale blue-shifted velocity structure preceded the onset of flare activity in that area by 120 minutes.

### 3.2.3 Predictions of Energetic Particle Emission of Solar Flares

The anticipation and recognition of the types and sizes of flares associated with energetic particles observed near the Earth are important aspects of solar activity predictions. Because there are extensive reviews (e.g., Svestka, 1976b) regarding this topic, we have restricted our discussion to a brief report of what can be determined about the particle emission of flares on the basis of the region's magnetic field configuration (also discussed in 3.2.1) and other flare emissions.

Energetic particle flares with few exceptions occur in magnetically complex regions and in particular in areas where the magnetic field gradients across the magnetic inversion line are strong. Severny et al. (1979) suggest that magnetic field gradients  $> 0.1 \text{ G km}^{-1}$  are necessary for proton flares to occur.

Flares producing energetic particles also generally have strong hard X-ray and microwave-radio bursts, as well as Type II and IV radio emission. McKenna-Lawlor (1979) finds that strong flare-associated microwave emission (indicative of particle emission) occurs in regions showing significant magnetic field changes and conversely, flares did not have strong microwave emission where no field changes were observed. Particle emission with flares can also be inferred by the occurrence of a filament eruption, Type II or Type IV radio emission. When these associated events occur it is likely that protons will be observed in space.

Homologous flares also provide a tool for anticipating the potential energetics of flares. Homologous flaring is observed in regions with magnetic field configurations conducive to flaring but which remains fairly static for periods of days. Under such circumstances the energy spectrum for subsequent flares (occurring anywhere from a few hours to several days) could be predicted on the basis of the original flare in the homologous set.

### 3.3 Solar Origin of Geomagnetic Activity

#### 3.3.1 Brief Historical Background

The association between observed flares and subsequent onset within one to three days of a distinct geomagnetic disturbance has been known and studied for more than a century. Solar astronomers and geophysicists since the days of Carrington and his well known observation of the white light flare of 1 September 1859 and subsequent great geomagnetic storm have been attempting to differentiate between the solar flares that are effective in the production of geomagnetic storms and those that are not. Originally only the optical (primarily H-alpha) observations of flares were available. With the passing of decades, additional information on flare emission ranging throughout the entire electromagnetic spectrum became available. This information included data on ionizing radiation as detected by ionospheric disturbances and/or X-ray flux, and radio frequency emission in both metric and centrimetric wavelengths.

#### 3.3.2 Two Types of Geomagnetic Disturbance, Both of Solar Origin.

Studies of geomagnetic storm data indicate two types of disturbances. The first, the so-called sporadic storms, occur primarily during the times of increasing and maximum solar activity. The second, the sequential or recurrent storms, occur more often during the declining phase of the solar cycle.

Sporadic storms are generally flare-associated and most often are preceded by a sudden commencement. The apparently causative flares can occur at all solar longitudes but most frequently are located in the central parts of the solar disk. The delay times between flare and storm onset can range from  $\sim 17$  hours to three days. The identification of flares that will be followed by geomagnetic storms and the estimate of the delay times constitute serious problems for those who forecast solar-geomagnetic activity.

Recurrent storms in sequential patterns ranging from 27 to 29 days occur throughout the entire range of the solar cycle. They are more numerous and more easily identified, however, during the decline phase of the cycle when the sporadic storms are less frequent. Chapman and Bartels, and Allen in earlier years attributed the source of these recurrent storms to so-called "M-regions" on the sun. Skylab observations (in 1973) led to the identification of these M-regions with low-emission regions in the corona, now called coronal holes. Coronal holes are regarded as the source of high speed solar wind streams which are currently considered to be the cause of the recurrent storms. Forecasters of this type of geomagnetic activity must be able to



identify the presence of a coronal hole on the solar disk and evaluate the potential for the development of a solar wind stream which may or may not become recurrent.

### 3.3.3 Recent Studies

In preparation for the International Solar-Terrestrial Prediction Workshop a number of papers were prepared which deal with problems of predicting both types of storms. These papers will appear elsewhere in the Solar-Terrestrial Predictions Proceedings.

A recent survey relating to some of these problems is a "Study of Geomagnetic Storms and Solar Flares in the Years of Increasing Solar Activity, Cycles 19 and 20 (1955-1957, 1965-1968)". This survey was prepared at the McMath-Hulbert Observatory by Dodson, Hedeman and Moehler and is published as Scientific Report of the Air Force Geophysics Laboratory, AFGL-TR-78-0267. During the years in question, 245 geomagnetic storms with maximum values of 3-hourly  $K_p$ 's  $\geq 5$  occurred. An evaluation of prior solar circumstances was made for each storm. Of the 245 storms in the seven years studied, 62% were considered to be flare-associated, 30% were primarily equatorial, and only 8% remained as problem cases. For the years studied, flares apparently were associated with a very high percentage of the most severe storms -- approximately with 100% for storms with maximum  $K_p$  as great as 8. See Figure 3.3. This same figure shows that recurrent storms, without appropriate prior flares, were associated with lower values of 3-hourly  $K_p$ .

In the years studied, flares that were found to have been followed by geomagnetic storms (maximum 3-hourly  $K_p \geq 5$ ) were, in general, important solar events. In evaluating prior flares, the usual optical evaluations of flares were supplemented by consideration of the Comprehensive Flare Index (See UAG-14). Fifty-four of the flares (36%) associated with geomagnetic storms were of H-alpha importance 3 or had a Comprehensive Flare Index  $\geq 11$ . For only 5 of the storm-associated flares was the H-alpha importance less than 2 and the Comprehensive Index also less than 6. The location of the flare on the solar disk apparently influences both the probability of storm-occurrence and its severity. See Figure 3.4.

Even though it has been shown that sporadic geomagnetic storms follow primarily important solar flares, only relatively small proportions of even "major" flares are followed by geomagnetic storms. ("Major" flares are flares that are far above average in at least one aspect of optical, ionizing, or radio frequency emission). An evaluation of the flare in terms of a Comprehensive Flare Index may be of some value to the forecaster. To assist in the prediction of flare-associated geomagnetic storms reference can be made to Figure 3.5. This figure shows that for the years 1965-1968, 269 "major" flares were identified. Only 58 (22%) of these flares were followed by geomagnetic storms. The graph also shows a systematic increase in the proportion of "major" flares followed by storms as the Comprehensive Flare Index increases.



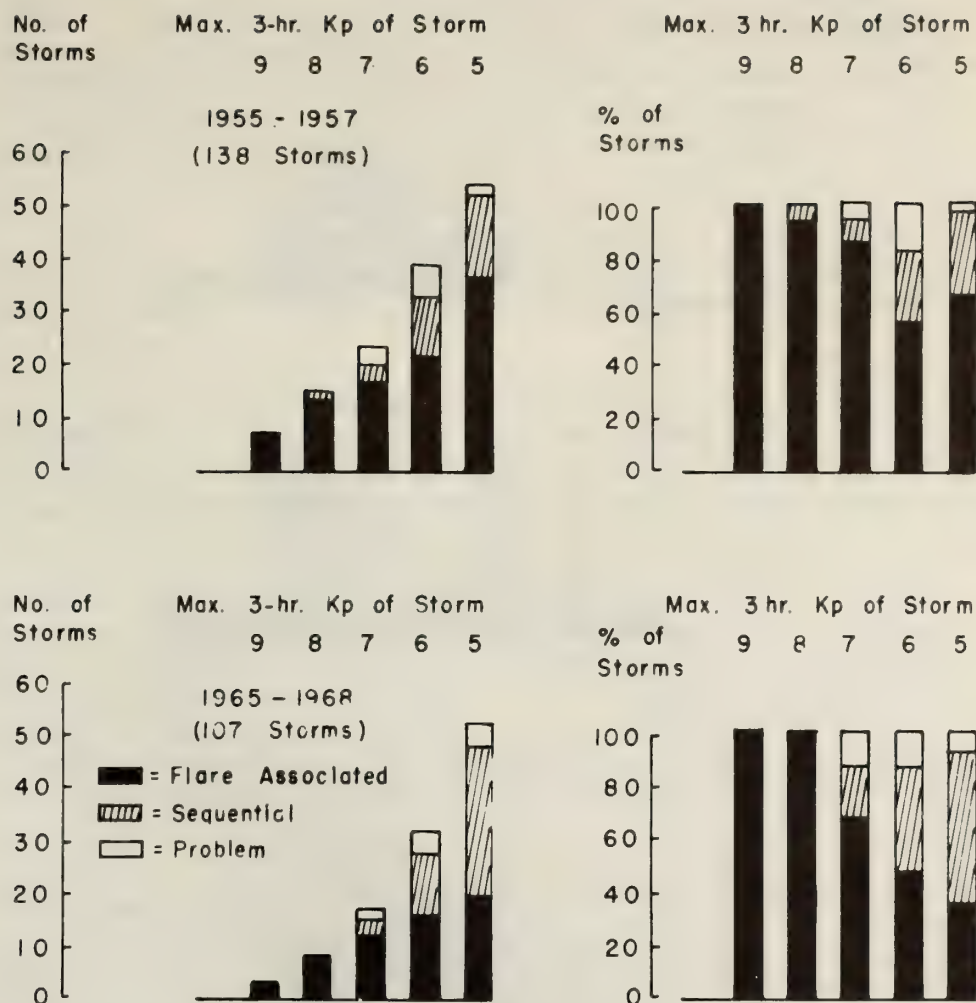


Figure 3.3

Number and percent of geomagnetic storms by severity and classification

#### 4. FORECAST TECHNIQUES

##### 4.1 Data Circulation and Exchange

Several categories of data are used in the preparation of forecasts. Each category of data supplies the forecaster with a small piece of information about the solar atmosphere, selected to satisfy relevant models developed around the evolution of active centers, and the place and time of flares, coronal holes (when they are reported,) geomagnetic disturbances, and sudden ionospheric perturbations. Cooperative efforts of observatories scattered around the world with various equipment on the ground and aboard spacecraft are needed to maintain a constant survey of the sun and to provide the data to forecasters. Once the data are acquired by an observatory, the problem then becomes one of circulation and exchange of that data both be-

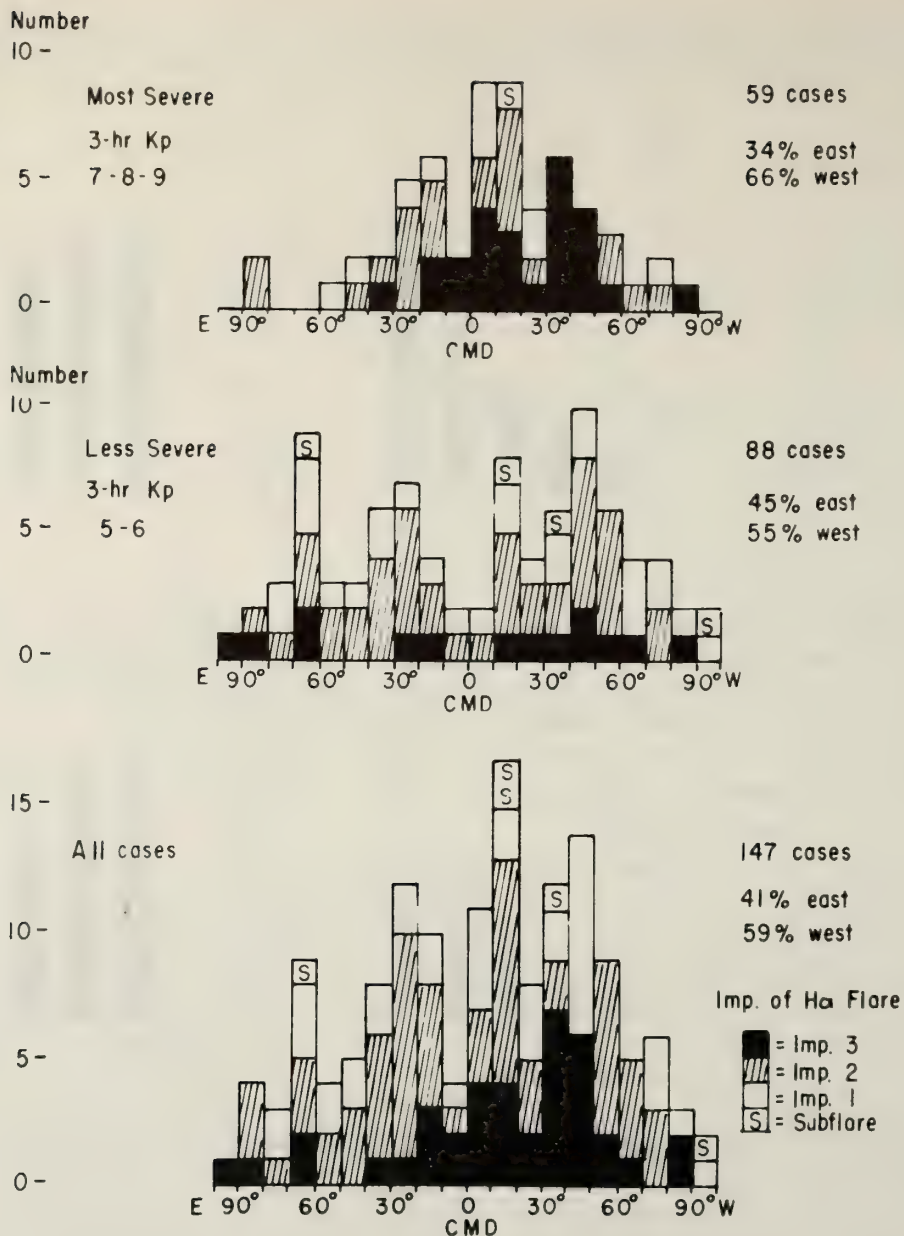


Figure 3.4

Central Meridian distance and importance of flares associated with geomagnetic storms of different severities; 1955-57 and 1965-68.

tween the observatory and the forecaster and between forecast centers, for all data types are not routinely monitored at the individual warning centers. This dual system of data exchange and circulation is dominated by the logistic and monetary restraints of data transmission.

Data are transmitted to the forecaster in plain language, codes, and pictorial formats. The forecaster depends upon the quality of the original data and the expertise of the observer. Reporting observatories usually transmit plain language reports in a common and familiar language and terminology; there are few problems in this method of data exchange. These plain language reports are very flexible, pointing out interesting features and trends, but due to their non-structured format they can omit important pieces of information. A coded format specifies in advance the parameters to be reported and the forecaster receives unambiguous numerical data. While pictorial data are expensive to transmit and some degradation occurs in the transmission process, this method of data circulation allows the forecaster to analyze the data personally.

The logistic and monetary restraints become more apparent when the world wide forecaster-to-forecaster data-exchange system is considered. The primary method of international data exchange currently in use is through coded messages. This is a relatively inexpensive method and language difficulties are minimized; however, this format is somewhat restrictive. The cost of world wide transmission of raw pictorial or digital data is prohibitive and would introduce problems of non-uniform data taking techniques (scale sizes, orientations, etc.). Very few plain language observations are circulated (internationally) between forecasters, which minimized potential language and terminology problems. All of these exchanges of data must occur on a daily basis and on an established schedule. If an observer cannot obtain data the forecaster must be made aware of this lack of data. In such cases the forecaster may be required to rely solely upon locally collected and analyzed data.

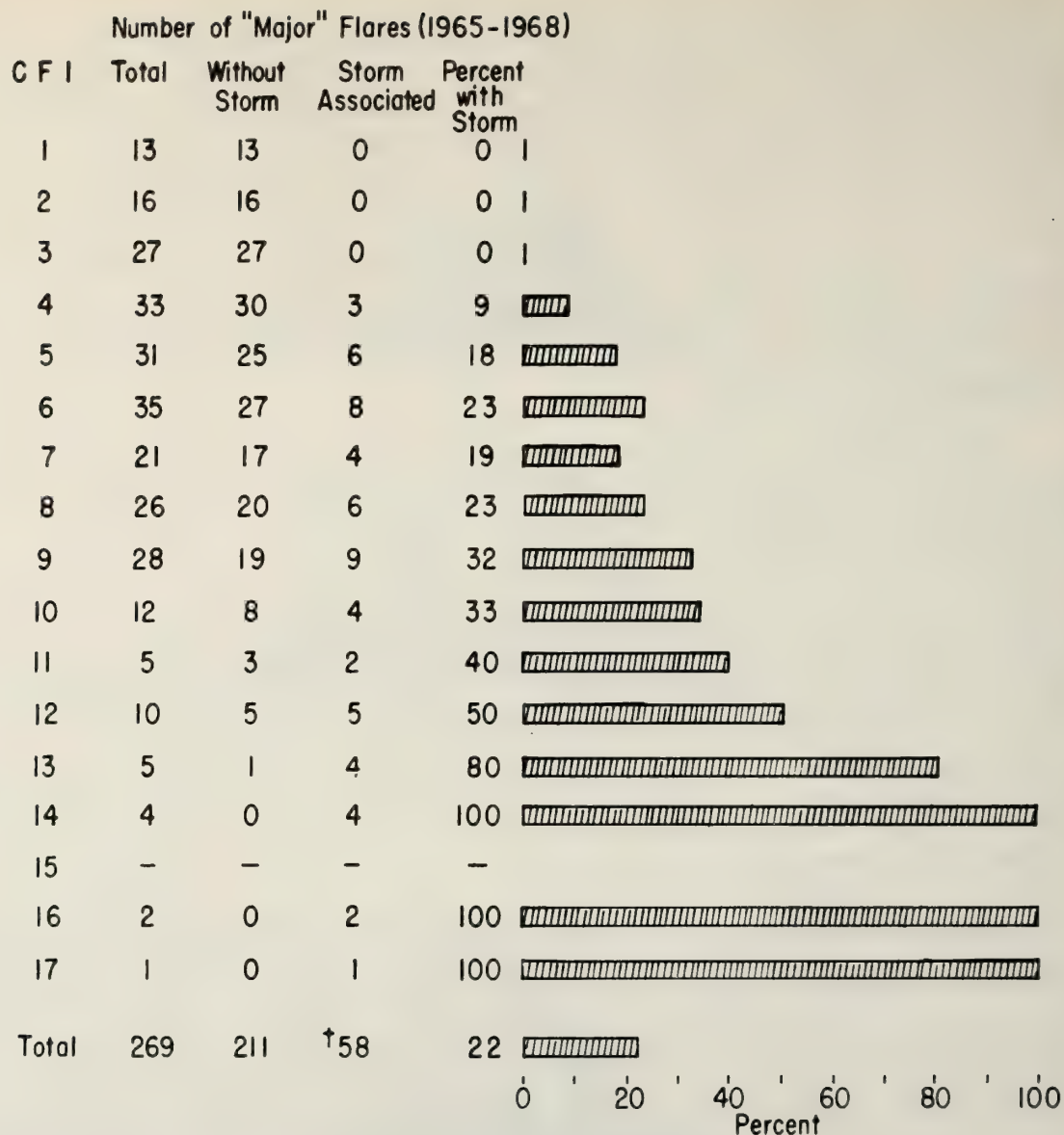
Event reporting is particularly difficult. Several categories of data are included in the reporting of a single event. Additionally, the same pieces of information are not reported if the delay in transmission varies. For example, in real time or near real time, an observatory could report the flux after a very short interval of time, or wait for the first peak and report the level and time of the peak or, with some delay, could summarize an event a short time after its end. There are two important difficulties: the identification of the event (is it really a flare or just a plage brightening) and the choice between the rapid relay of information and its precision (position of a flare or evolution of its brightness).

In conclusion, it should be emphasized that the data circulation and exchange is an important consideration in solar forecasting. Its organization depends upon several choices: quality of equipment, expertise of the staff members, technique of transmission (pictorial, plain language or coded messages), stations contributing to a national network but established in foreign countries, and the quality of the sites and their number (the number of stations needed in order to obtain a permanent optical survey is closely linked to the mean quality of the sites and to their latitude.)

#### 4.2 Human Analysis of Data

The forecasters are facing several problems: (a) identification of active regions, (b) understanding the coronal structure, (c) evaluation of links between active regions, (d) forecasting the evolution of all the ac-





†Includes the "ambiguous" as well as the uniquely storm-associated flares.

Figure 5

Number and percent of "major" flares  
with different values of the comprehensive flare index (FFI),  
associated with geomagnetic storms with maximum 3-hourly  $K_p \geq 5$ , 1965-1968

tive regions, and (e) forecasting the flaring results of this evolution. In addition, most of the data are not original but are telecopied or are reports that are in plain language or in codes, and finally several categories of data or reports can be missing.

Active regions can be identified by the photospheric field, the calcium plages and the sunspot evolution. Information on the upper structure of active centers is available in H-alpha data (arch, filaments, prominences, loops, etc.), in UV and X-ray images, in radio heliograms and in coronal data

at the limb. Global radio flux and X-ray emission can be used to evaluate events. The links between adjacent active regions are related to their distance from each other and to their relative phase of evolution. The forecaster must take into account the rapidity and the type of this evolution, the links, the results of the differences in latitudes and of the phase of the solar cycle. The sixties have seen the new contributions of radio astronomy and photospheric magnetograms, while the seventies have produced the first results of UV and X-ray imaging. However, active regions are not fully understood and, despite recent progress, our knowledge of the flare process is still rather poor.

This lack of a physical model could explain the current reports of the forecasting centers. Their forecast is primarily based upon a series of criteria which were established by studies of correlation between several series of specific data and the occurrence of events. Most of these criteria are related to the photospheric magnetic field of active regions, either longitudinal components of actual magnetic field observations or "inferred" magnetic fields resulting from other observations such as H-alpha spectroheliograms, comparison of H-alpha and Ca spectroheliograms and sunspot data. In addition, they take into account the microwave emission (10 cm or 3 cm wavelength), the proper motions of sunspots, the complexity of active regions, the recent flaring activity, the "active longitudes" and the rapidity of evolution. At the end, the forecast is issued in terms of flaring activity of each active center expected on the next day, or of global flaring activity of the solar disk on the next three days. Sometimes the conclusion is as follows: Much of the forecasting is dependent on the expertise of the forecasters, emphasizing that the forecasters are aware of the weakness in their forecasting method(s).

#### 4.2.1 Active Regions

A "simple and isolated" active region is bipolar. During its evolution only a few subflares will occur near the beginning of its growing phase. The neutral line is nearly parallel to a meridian, the coronal contribution is weak in terms of 10 cm and X-ray emission, and the plage and filament will disappear after one or two rotations. These centers are born in areas of the Sun that are free of remnant magnetic fields. The first step of a forecast is the selection of these "quiet" regions, which may account for 80% or more of the active regions on the disk.

Any other active region is "anomalous". This "anomaly" usually results from a link with another active region. The simplest anomaly is the Bf active region which emerges on the West side of another region; its trailer spot is larger than its leader. Another anomaly is the "parasitic polarity" or the "satellite spot", which consists of a small area of opposite unipolar polarity inside the unipolar part of a plage. The largest anomaly results from the growth of an active region inside an old plage or imbricated with another active region; we then obtain a "complex" active region. The forecast of the evolution of these active regions requires a careful inspection of all the data. Reversed polarity (from the normal for the hemisphere and the cycle), east-west orientation of the neutral line, a convoluted neutral line (intrusions of opposite polarity, etc. are anomalous conditions, and most flares occur in "anomalous" active regions. An important item in the forecast is the



sense of evolution of these regions, i.e., whether they are growing or decaying. During the growing phase, the region is compact and the sunspots, even the large ones, are frequently not seen through the enhanced chromosphere. Roughly speaking, during the decay, an active region becomes simpler and the distance between the opposite polarity spots increases. The process of simplification can modify the neutral line and is accompanied by flaring activity.

The microwave brightness (near 3 cm wavelength) of the active region is linked to the intensity and to the gradient of the magnetic field. The coronal contribution in terms of 10 cm flux and X-ray emission can be very high. Another important item of the forecast is the latitudinal extent of the group and its actual latitude. For latitudinal extents less than  $10^0$ , the differential rotation may play a role in the motion of the spots.

In the forecaster's study of anomalous active regions, all the data must be used in a "cross-checking" process. A strong gradient of the magnetic field is not always easy to identify just according to a crude real time magnetogram. The microwave and X-ray emission could confirm this identification. In the first step of the forecast, all these data are not used as criteria, but as pieces of information in a jigsaw puzzle.

#### 4.2.2 Flares

Solar flares occur during a specific evolution of the magnetic field near the inversion line of the longitudinal component of the field. For example, if the two polarities evolve in opposite ways (one growing, the other declining), a flare is triggered although the time of the flare cannot be predicted. There are at least two categories of flares: the "two ribbon" flares and all others, where the latter form the largest group of flares.

A series of magnetograms with high spatial resolution ( $\sim 2$  arc sec.) shows the temporal evolution of the magnetic field inside each unipolar part of a plage. The intensity of the field does not vary in a monotonous way from one border of this area to the other. One can identify a series of cells of various sizes, with a peak of field intensity inside each cell. Each cell evolves separately, at random (the evolving magnetic features). This evolution could explain why spots appear as separate magnetic cells and why many spots on a time scale of 24 hours, appear and disappear at random. In the cells, the spots are the peaks of the field.

This incoherent evolution of unipolar fields explains many of the sub-flares in complex active regions. The polarities are complexly mixed, and there may be a large number of magnetic cells of opposite polarity separated by a convoluted polarity-inversion line. Their random evolution gives rise to a higher probability of increasing complexity than in a simple bipolar group in which only a few cells of each polarity are separated by a simple inversivity line. Consequently, the forecaster can identify easily the active regions in which small flares will occur. Sometimes one can specify the sites of the flares (e.g. satellite spots) or the area in which flares will occur, and can predict that several small flares will occur during the next 24 hours. However, we cannot specify the time of the flares or the actual site of many of them.



The large "two ribbon" flares occur along a long inversion line. Unfortunately the process of "flare build up" seems to operate with varying speeds. Sometimes the flare occurs promptly after the formation of this line, or several days may pass before the flare occurrence. Of course, sometimes the flare does not occur, at least on the visible side of the sun. New hypotheses suggest that many categories of evolution of the magnetic field could contribute to a "destabilization" of the prominence which may trigger the flare. Among them are the evolution of the active region itself, the appearance of emerging flux and the birth of a new active region both near the existing "flare build-up" region. Using the various reported data, the active region "flare build-up" can be predicted, and may be verified by the microwave and the X-ray emission, but, again, the time of occurrence is not predictable. Two large flares may occur on the same day; several days may elapse between the major flares. In addition, according to the observed or inferred gradient of the magnetic field, a proton event can be forecast. Of course, the proton flux will not always reach the Earth environment, but the hard X-ray burst and the strong microwave burst attest to the presence of the acceleration process.

Of course, not all flares can be placed in these two categories; many do not correspond to either description. However, these two models help in preparing the forecast and in illustrating the difficulty of obtaining a high score in forecasting large flares. The delay between the first forecast of the large flare and the flare occurrence can reach several days, increasing the reluctance to maintain the forecast and leading to termination, of the forecast, only to have the flare occur later.

Summarizing this forecasting technique, the first step is to forecast the evolution of all active regions, taking into account their history, the position in the solar cycle, and the latitude effects. In this effort, all the categories of data are used to confirm the evaluation in terms of a coherent active region model. The second step concerns the flare forecast itself in which several types of data supply us with criteria for prediction of the type and frequency of expected flares. However, data availability at forecast centers varies and depends partly on weather conditions. During the fall and winter months, there is sometimes very poor optical coverage, and for several consecutive days (or weeks) there may be no optical observations at the forecast center(s) and the forecasts may be based only on radio and satellite data and on encoded and plain language reports.

#### 4.3 Statistical and Computer Assisted Techniques

A wide variety of solar observations has become available on which to base solar flare forecasts. The perplexity in the use of the resulting numerous parameters in deriving prediction criteria leads, quite naturally, to the desire to use systematic techniques for extracting predictive information. In this section we consider recent work in statistical and computer-assisted forecasting techniques, including their performance, impact, and possible limitations. The aspect in common to all these methods is an objective formula or procedure, derived from historical data, which is then applied to future time. Although the rules for evaluating forecasts apply

equally well to any prediction method, we will limit the discussion here to those methods using many input parameters and which exclusively predict the occurrence of flares.

#### 4.3.1 Methods and Performance

A number of contributions to these proceedings use techniques such as multiple regression, pattern recognition, discriminant analysis, and logistic regression. A mathematical presentation of these is beyond the scope of the discussion here. Instead, we will discuss the forecast accuracies, taking particular note of the non-uniformity of verification procedures used in the various studies.

The obtained accuracy of a forecast is severely affected by the following factors: a) the level of solar activity during the test period, b) the range in magnitude of the activity predicted (for example, flares taken of all sizes as opposed to a smaller group of some given class), c) the interval of time for which the forecast is made, and d) biases due to the selection of the more flare-productive regions. The probability of an event occurring within a given "window" obviously can be increased by widening the ranges of time and magnitude or by increasing the level of solar activity. In the case of the latter we need to note if there has been any pre-selection of the data which leads to an effective increase of activity due to bias toward the more active regions on the sun. It is permissible to make such selections, of course, but in any case the final climatological probabilities for the test period should be clearly stated. Otherwise it is not possible to make even a qualitative comparison of the various methods. Table 1 summarizes the various contributed efforts, where a "Yes" in the column labeled "Selected Regions" indicates the use of data which may be biased toward increased activity. Forecast accuracies attained by a conventional (non-objective) forecast obtained from the Space Environment Services Center (SESC) in Boulder are included for a comparison. All forecast accuracies are given in terms of percentage of correct forecasts and refer to predictions made for individual active regions except as noted otherwise.

In addition to the studies listed above we may also note some recent preliminary results from the National Bureau of Standards (Vecchia, 1979, private communication), where the logistic regression technique obtained accuracies of 89 and 83% for no-flare and flare X-ray class C, respectively. The corresponding climatological probabilities were 80 and 20%, using a data base limited to active regions containing sunspots.

The relative success of the various forecasts could, in principle, be roughly determined by comparing the ratios of the forecast accuracies to the climatological probabilities. However, the lack of sufficient statistics and uniform data bases make a strict comparison impossible. It is obviously preferable to test all the various methods on a common base of solar data. In any case the results of Table 1 are encouraging enough to indicate that objective techniques may already be useful in solar activity forecasting.



While the techniques described here remove subjectivity in the preparation of the forecast, they in no way eliminate the subjectivity in many of the descriptive or morphological input parameters. For example, no scheme is presently available to quantify the impressions that a forecaster may derive from visual inspection of a hydrogen-alpha photograph. In any case the use of objective forecasting procedures will shift the emphasis more strongly toward the acquisition of solar data, which will now be required to have greater consistency than ever before.

#### 4.3.2 Limitations in Forecast Accuracies

We now consider the possibility that the accuracy of any forecast, using the presently available solar data, is fundamentally limited by Poisson statistics. If the solar data consist exclusively of parameters which, at most, are only statistically associated with flare incidence (which is probably the case since no viable flare precursors are yet available, nor is the flare mechanism known), and if the parameters vary with a time-scale on the order of one day (which is effectively true since the data are collected daily), then it is hypothesized that the maximum information obtainable for a daily forecast will not exceed that which is obtained from a precise knowledge of the mean daily rate of occurrence of the phenomenon to be predicted. In this case, the prediction accuracy will be limited by the random arrival of events within the forecast interval. For example, if the predicted event has a perfectly known mean rate of occurrence of two per day (typical of small, geophysically insignificant flares) then, according to the Poisson distribution, the probability that one or more such events will occur on a given day is about 0.84; in this case forecasts for "flare" would be limited to about 84% accuracy. On the other hand, if the mean rate were 0.25 per day (as in the case of larger, geophysically important flares) the success rate would only be 20%. It would actually be better, from the standpoint of forecast accuracy, to always predict "no flare" for the latter situation; but of course the forecasts would then have no utility. This dilemma emphasizes the need for the continued search for the triggering mechanism of solar flares.

#### 4.4 Forecast Users and Format

Since solar flares cannot at this time be accurately predicted, several forecast formats have evolved in an effort to satisfy the majority of the user community. One of the major hurdles in the standardization of a forecast format is that the requirements of the users are as varied as their disciplines and the solar effects upon their systems.

There exist several types of forecast formats of which the probability and the yes/no formats comprise the majority. There are several advantages and disadvantages to each method, and their applications to different users sometimes present potential problems. Of course, the ultimate goal of all warning centers is the ability to produce a completely accurate yes/no forecast.

A warning center may use a combination of probability and yes/no forecasts, such as a percentage probability for the occurrence of an X-ray flare capable of producing an SID or a yes/no forecast on whether or not a particu-



T A B L E 1

AUTHOR(S)	METHOD	FORECAST INTERVAL (DAYS)	SELECTED REGIONS	PREDICTAND	CLIMATOLOGICAL PROBABILITY (%)	FORECAST ACCURACY (%)
Burov et.al.	Pattern Recognition	1	No	No Flare Flare $\geq C$	88* 12*	96 52
Gumanitsky et.al.	Image Identification	2	Yes	Not Flare Subflare and importance 1 Importance 2	66 30 4	91 52 36
Purple Mountain Observatory	Empirical Probability	3	No		?	77**
Purple Mountain Observatory	Empirical Probability	3	No	No Proton Event Proton Event ( $\geq 10 \text{ cm}^{-2} \text{ s}^{-1} \text{ stran}^{-1}$ )	99.4 0.6	99.9*** 12.5***
	Multiple Regression	3	No	No Proton Event Proton Event	99.4 0.6	99.9*** 7 ***
Jakimiec and Wasiucioneck	Multiple	1	Yes	No Flare Importance 1	73 27	87 68
Hirman et.al.	Discriminant	1	No	No Flare Flare $\geq C$	88 12	95 63
SESC Aug-Dec 1977		1	No	No Flare Flare $\geq C$	88 12	95 62

\* Climatological probability derived from NOAA data used in the analysis

\*\* Overall accuracy including all classes

\*\*\* Full Disk probability

lar flare will produce protons which intercept the earth. All forecasts can be issued in a yes/no format, as is the practice at some warning centers. The confidence in this format can be modulated by such qualifying words as "caution" and "doubtful". In some cases the user requirements define the forecast format. These instances are usually resolved on an individual user-to-warning-center basis. No matter which format is used, the education of the user into the interpretation, limitations and strengths of the forecast is imperative.

#### 4.4.1 Forecast for Customers using the Forecast to Begin or Continue a Program

Their decision is taken in the framework of their entire project: their problem is to know if they have to take a chance now or to wait for the next event of the same category. Of course they would like to know the degree of confidence that they can put in the forecast. Obviously if the forecaster has some specific reason for having low confidence in his forecast, he would not issue it. On the other hand, most of the time, the user cannot evaluate in advance the reliability of the forecast without the forecaster's evaluation.

The true problem for the users conducting scientific experiments is to know if they will have future chances of observing the same category of event before the end of their enterprise either owing to the actual trend of the event occurrence or to the time degradation of their equipment. It raises a question of the utility of a mean range forecast showing the probability,

in this part of the solar cycle, of having various categories of events: one proton flare each forty days, etc... The reliability of more specific "mean range" forecasts can be questioned except for the forecasts of recurrent geomagnetic activity at times in the solar cycle when recurrent geomagnetic storms dominate.

#### 4.4.2 Customers Using Forecasts as a Warning.

A forecast helps them to promptly identify the source of an unusual disturbance: S.I.D., strong magnetic storm, etc. However, these unusual disturbances rarely occur. In order to observe them, it is necessary to have a large event occurring at the right time of the day. Consequently the forecast reliability may only be very poor. It seems that a three "color" code as used by the USAF would be the simplest format for these customers, with the understanding that the event is not expected, possible, or expected.

#### 4.3.3 Customers using Forecasts to Issue Their Own Forecast.

These customers require a very accurate forecast: the ionospheric forecasters need to know the time of the geomagnetic storm because its effect depends upon the time of occurrence during the day. The satellite orbit forecasters require as accurate a solar flux level prediction as is possible. Finally, after this careful review of the specific needs of the customers, it is understandable that the format must be adjusted to the various requests of the customers. In order to maintain confidence, forecasts must be based on a reasonable scientific background.

### 4.5 Staff Problems

The reliability of forecasts all too often depends largely on the experience or "expertise" of the individual forecaster or small groups of forecasters working together. Obviously, the forecasters with the most expertise cannot be available to make every forecast. Since various forecasters may use their own methods, the net result can be a loss of reliability and homogeneity.

Basically, it is desirable to employ scientific forecasting techniques, which require more understanding than "feeling" on the part of the forecaster. Therefore, a link between the forecaster and the solar scientific community must be maintained. In some instances the forecaster is a scientist; in others, a scientist provides guidance. A scientist acting in the latter role must be interested in and aware of all the problems involved in forecasting. One advantage of this arrangement is that it should require less training to bring a forecaster to a certain level of skill through understanding than by attaining the "expertise" through experience. This also helps lessen the impact when a high rate of exchange in personnel is involved.

The staff problems may eventually be partially resolved by using computers to issue the forecast. More inputs need to be determined and refined before we can expect successful operation on a daily basis, but multivariant discriminate analysis techniques, for example, have shown some encouraging



skill. Automation could therefore modify the forecaster's role to that of an analyst selecting or verifying inputs, or devising and testing new prediction algorithms, or tailoring forecasts to the needs of particular users.

## 5. CHECKING THE FORECAST

### 5.1 Probability of Success

Only the forecasts issued in a standard, quantitative format can be objectively evaluated. Plain language, subjective descriptions of the active regions and their flare potential are helpful in informing the forecast users (and other centers), but their accuracy cannot be objectively verified.

The 10 cm flux and A indices forecasts are easy to compare with observed conditions and one can compute the scatter of the data. In this field a three day forecast is generally issued. This kind of forecast is easily done since numbers are predicted but it has been shown how promptly the reliability decreases with time. The third day forecasts are frequently very poor.

Checking the probability of success of the flare forecast is not an easy task. Whatever the forecast format may be, there are always both the class "Quiet" and "No Flare" and the class "Eruptive" or "Flare". At the time scale of one day (a daily forecast), only a few active centers are flaring, especially if we are interested in large flares. Consequently, if we only take into account the success of the "No Flare" and "Flare" forecasts, it is easy to produce 80 or 90% correct forecasts, even if no large flares are successfully forecast. Papers in which new forecasting methods are suggested claim up to 90 or 95% successful flare forecasts because of this scarceness of large solar events. The comparison must only be established between "positive forecasts" and relevant events: the scarcer the events, the poorer the reliability. In addition to this kind of test on a long interval series of forecasts, we can select shorter intervals of time during which all the data were available in order to evaluate the influence of a lack of data upon the result.

### 5.2 Utility for Users

The users must know the general trend relevant to the category of events they want to observe and the current probability of success of the relevant forecast. They must use this information in the planning of their experiment. Sometimes users have a list of independent conditions which must be fulfilled together for the success of their experiment but may not consider that any condition not easily fulfilled will strongly decrease the probability of carrying out the experiment. If, in addition, the forecast has a low probability of success, they may never perform their enterprise.

However, they must also be shown the utility of following a forecast and, in this case, the scarcer the events, the more useful the forecasts. Such evaluations can be easily derived by comparing the probability of success of the forecast to the daily probability of occurrence of the event during the period of the experiment; these are frequently of the order of 10 to 50.



### 5.3 Quantitative and Qualitative Forecasts

By qualitative forecast, we mean the use of crude categories of events, and by quantitative forecast, the use of probability. A check of the qualitative forecasts shows the difficulty of selecting the categories of events. As an example, during four consecutive days, Meudon issued the forecast "Active" for an active region where a large flare with large microwave burst occurred on the fourth day: it was considered a true success because the evolution of this category of centers is slow and at the present time no one can identify any difference in the active region features on three, two or one days before the major flare. However, statistically, there will be just a 25% of success. In addition, during this long flare waiting interval, many small subflares occurred in a corner of the plage, near a "satellite spot". These events were not forecast because they were small by comparison to the expected large flare, but they will appear in the statistics as not forecast. Probably, the best way to check the scientific value of a forecasting method would be to compare the number of successfully forecasted events to the number of reported events.

Checking the quantitative forecast is not an easy task of the need for as many boxes as groups of probability. One can compute the actual probability but it would be very difficult to understand why the forecast was a failure. Of course, one can carry out a multiregression analysis but there is frequently no significant difference between the components.

## 6. CONCLUSIONS AND RECOMMENDATIONS FOR IMPROVEMENT IN THE SHORT TERM FORECASTING OF SOLAR ACTIVITY

Short term prediction of solar activity is a basically difficult undertaking and currently is imperfect. There is, however, hope and belief that improvement can be expected in the years and decades ahead. The primary recommendations are: 1) improvement in the scientific basis of the forecast (with assured financial support), and 2) increased ability to forecast the time of flare occurrence.

The specific recommendations are outlined below. The recommendations are presented according to present, near future and long-term future situations.

### 6.1 Forecasting of Solar Activity

#### 6.1.1 Present Needs and Available Material

- A. Permanent survey of solar activity in H-alpha, microwaves, and X-rays.
- B. Prompt access to data (daily).
- C. Currently available data (must continue to be provided):
  - (a) Optical data: H-alpha, Ca II, white light, (b) high resolution magnetic and velocity field observations, (c) radio heliographic data --

microwave and metric wavelength, (d) soft X-ray global flux (real time), (e) radio global flux at fixed frequencies, (f) helium I (10830) spectroheliograms (coronal holes) and (g) white light coronal data (coronal holes).

#### 6.1.2 Near Future Requirements (a few years hence)

- A. More frequent reports of pertinent data.
- B. Magnetic field data with high spatial resolution ( $\geq 2''$ ) and report of rapid evolution.
- C. Velocity field data of high spatial resolution.
- D. Vector magnetic field data.

#### 6.1.3 Long Term Future Needs (many years or decades hence)

- A. Far-side view of sun.
- B. X-ray images.
- C. EUV images.
- D. Global magnetic dynamo models.

### 6.2 Diagnosis and Forecasting of Flares

#### 6.2.1 Present Needs and Available Material

- A. Permanent survey of solar activity in H-alpha, microwaves, and X-rays.
- B. Prompt access to the data (daily).
- C. Currently available data (must continue to be provided):
  - (a) optical reports (H-alpha), (b) emerging flux, (c) microwave brightness of active regions, (d) magnetic field strength and gradients in active regions, (e) photospheric and chromospheric velocity field patterns, (f) radio spectrographic and polarization data, (g) global soft X-ray flux, (h) X-ray burst reports, and (i) particle monitoring.
- D. Improved flare evaluation through more general derivation and use of a comprehensive flare index based on ionizing and radio frequency emissions as well as on the optical manifestations of the flare.

#### 6.2.2 Near Future Requirements (a few years hence)

- A. Working flare models.
- B. Pre-flare information:

(a) H-alpha precursor, (b) filament activation (c) X-ray precursors, and (d) soft X-ray images.

C. Improved magnetic and velocity field data of high spatial resolution and information concerning magnetic shear.

D. Data from Solar Maximum Mission (SMM):

(a) Hard X-ray images and (b) chromospheric polarimetry.

E. Fine structure of spot region evolution (high time and spatial resolution in white light).

F. High spectral and temporal resolution microwave intensity images, with polarization.

G. Verification of successful use of current parameters now used in flare forecasting.

H. Wave propagation (oscillation and fine structure of microwave emission).

### 6.3.3 Long Term Future Needs (many years or decades hence)

A. Derivation and use of quantitative parameters for flare-forecasting.

B. Improved working flare models.

C. Improved active center models.

The location of the flare on the solar disk apparently influences both the probability of storm-occurrence and its severity. See Figure 3.4.

Even though it has been shown that sporadic geomagnetic storms follow primarily important solar flares, only relatively small proportions of even "major" flares are followed by geomagnetic storms. ("Major flares are flares that are far above average in at least one aspect of optical, ionizing, or radio frequency emission). An evaluation of the flare in terms of a Comprehensive Flare Index may be of some value to the forecaster. To assist in the prediction of flare-associated geomagnetic storms reference can be made to Figure 3.5. This figure shows that for the years 1965-1968, 269 "major" flares were identified. Only 58 (22%) of these flares were followed by geomagnetic storms. The graph also shows a systematic increase in the proportion of "major" flares followed by storms as the Comprehensive Flare Index increases.

## 7. Annex 1, Sheared Magnetic Fields

There is a considerable variety of solar active region phenomena which may be attributed either to sheared magnetic fields or to the process of shearing. We will briefly describe the physical process and its potential for storing flare energy, as well as the observable effects which may be useful in forecasting operations.



By a shear we mean the lateral displacement of magnetic field footpoints, which, in a conducting plasma, induces electric currents to flow along the field lines connecting the footpoints. This current produces a twist in the field lines, which, in turn, may then influence the chromospheric and photospheric structures, resulting in patterns which are identifiable in optical observations.

From the standpoint of flare research the most interesting aspect of sheared fields is their energy-storing capability. Energies in excess of  $10^{32}$  erg can easily be contained in sheared, thousand-gauss fields with scale sizes typical of active regions. Since this energy exceeds that contained in a magnetic potential configuration, it may be extractable when the sheared fields are subjected to an instability. Newly-emerging magnetic flux and critical shearing angles have been suggested as possible triggers for the sheared field relaxation. It is an obvious necessity to discover what conditions and critical limits are required to initiate the flare energy release; it is an additional benefit with the case of sheared fields that the magnitude of the flare, when it does occur, could in principle be determined in advance by monitoring the accumulation of energy during the pre-flare shearing process.

We now list the solar phenomena associated with shearing, the types of activity reported to accompany them, and the methods of observation. Again, we emphasize that the association of these phenomena with flares (as well as the current use of several in forecasting schemes) may be understood on the common basis of magnetic shear.

<u>Phenomenon</u>	<u>Associated Type of Flare</u>	<u>Method of Observation and Observational Requirement</u>
1. Sunspot Motions (Translation)	All	White light or near continuum; one day base preferable
2a. Sunspot Rotations	All	White light or near continuum; one day base preferable
2b. Spiral structures in sunspot penumbra*	Proton	White light, instantaneous
2c. Spiral structures in chromospheric superpenumbra	All	H-alpha; instantaneous
3. Fibril inclinations near neutral line	All (includes spotless flares)	H-alpha; instantaneous; requires knowledge of geometry in the minimum energy state for comparison
4. Existence of neutral line filament	All (includes spotless flares)	H-alpha; instantaneous
5. Velocity Shear (Direct)	All (includes spotless flares)	Dopplergram; instantaneous; requires region to be $\geq 30^\circ$ from disc center

6. Magnetic Shear (Direct)	All	Vector magnetogram; instantaneous
7. Displacement of ribbons in smaller precursor flares	Large	H-alpha; one day base, or longer

\*It is presumed that the largest shears will be noted even in penumbrae, hence the association of the spiral structures with the largest flares.

## 8. WORKING GROUP MEMBERS

Various sections of this report were drafted separately by a few individual members of the Working Group. These sections were then reviewed by several Working Group members, whose suggestions were then incorporated. The report therefore reflects the general ideas and suggestions of those who attended the meetings of the Working Group on Short-Term Predictions of Solar Activity. Unanimity was not total but the members of the Working Group who were present at the Work Shop were in general agreement on the ideas expressed in the final report and on the recommendations made by the Working Group. The following representative members of Working Group A2 attended the Work Shop:

Paul Simon (Chairman)	Meudon Observatoire, Meudon, France
Jesse B. Smith, Jr. (Co-chairman)	NOAA/SEL, Assigned NASA/MSFC, AL, USA
Ding You-ji	Yunnan Observatory, China
W. Flowers	NOAA/SESC, Boulder, CO., USA
Guo Quanshi	Purple Mountain Observatory, China
Karen L. Harvey	Solar Physics Research Corp., Tuscon, AZ USA
Ruth Hedeman	McMath-Hulbert Observatory, Pontiac, MI, USA
Sara F. Martin	Calif. State Univ. at Northridge (San Fernando Observatory), CA, USA
Susan McKenna Lawlor	St. Patrick's College, Maynooth, Ireland
Lin Yuam-zhang	Peking Observatory, China
Donald Neidig	AFGL/Sacramento Peak Observatory, NM, USA
V. N. Obridko	Izmiran, Moscow, USSR
Helen Dodson Prince	McMath-Hulbert Observatory, Pontiac, MI, USA
David Rust	American Science & Engineering, Cambridge, MA, USA
D. Speich	NOAA/SESC, Boulder, CO USA
Arnold Starr	Holloman AFB, NM, USA
N. N. Stepanyan	Crimean Observatory, USSR

## 9. REFERENCES

- Ding You-ji, Luo Bao-rong, Zhang Bai-rong, Li Wei-Bao (1979a): Review of proton flare prediction by means of spiral spots in 1978. Preprint.
- Ding You-ji and Zhang Bai-rong (1979b): Spiral sunspots and solar activity forecasts. P-STPW.
- Dodson, H. and Hedeman, E. (1970): Major H- flares in centers of activity with very small or no spots. Solar Phys., 13:401.
- Dodson, H. W., Hedeman, E., and Mohler, O. C. (1979): Examples of 'problem' flares or situations in past solar-terrestrial observations. P-STPW.
- Giovanelli, R. (1939): The relations between eruptions and sunspots. Astrophys. J., 89:555.
- Harvey, K. and Harvey, J. (1979): Photospheric velocity fields as indicators of flare activity. P-STPW.
- Harvey, K. L. and Martin, S. F. (1979): Forecasting of solar flares based on magnetic field configurations. P-STPW.
- Harvey, J. W. and Sheeley, N. R., Jr. (1979): Coronal holes and solar magnetic fields. Space Science Reviews, in press.
- Heyvaerts, J. (1974): Coronal electric currents produced by photospheric motions. Solar Phys., 38:419.
- Marsh, K. (1978): Ephemeral regions flares and the diffusion of network. Solar Phys., 59:105.
- Martin, S. F. (1979): Preflare conditions, changes and events, a review of observations. Solar Phys., in press.
- Martres, J. -J. (1968): Originedes regions ab actives solares 'anormales', in IAU Symposium No. 35, p. 25.
- Martres, M.-J., Soru-Escaut, I., and Rayrole, J. (1973): Relationship between some photospheric motions and the evolution of active centres. Solar Phys., 32:365.
- Martres, M.-J., Rayrole, J., Ribes, E., Semel, M., and Soru-Escaut, I. (1974): On the importance of photospheric velocities in relation to flares. In Flare-Related Magnetic Field Dynamics Conference, Boulder, Colorado, p. 333.
- Martres, M.-J. and Soru-Escaut, I. (1977): The relation of flares to 'Newly Emerging Flux' and 'Evolving Magnetic Features'. Solar Phys., 53:225.
- McIntosh, P. S. (1979) Private Communication.



- McKenna-Lawlor, S. M. P. (1979): Short-term prediction of the potential of an active region to produce recurrent proton flares. P-STPW.
- Lang, K. R. (1979): Solar flare prediction using radio wavelength interferometers. P-STPW.
- Neidig, D. F. (1978): High resolution observations of fibril changes in a small flare. Preprint.
- Neidig, D. F., DeMastus, D. L., and Wiborg, P. H. (1978): Flares, force-free fields, emerging flux and other phenomena in McMath 14943. Report AFGL-TR-0194.
- Priest, E. R. and Raadu, M. A. (1975): Preflare current sheets in the solar atmosphere. *Solar Phys.*, 43:177.
- Rust, D. M. (1976a): Flares, observations of flare-associated magnetic field changes. *Phil. Tran. Roy. Soc. London.* A281:427.
- Rust, D. M. (1976b): An active role for magnetic fields in solar flares. *Solar Phys.*, 47:21.
- Rust, D. M. (1979): Transient coronal holes and magnetic reconnection. Submitted to *Solar Phys.*
- Severny, A. B., Stepanyan, N. N., and Steshenko, N. V. (1979): Short-term forecast of evolution and flare activity of active regions in Soviet Works. P-STPW.
- Skylab Workshop on Flares (1979): Solar flares, monograph from the Skylab Workshop on flares.
- Smith, J. B. (1972): Predicting activity levels for specific locations within active regions. In Solar Activity Observation and Predictions, (eds. P. S. McIntosh and M. Dryer), p. 429.
- Sturrock, P. (1972): Magnetic models of solar flares. In Solar Activity Observations and Predictions, (eds. P. S. McIntosh and M. Dryer), p. 163.
- Svestka, Z. (1976a): *Physics of Solar Planetary Environments*, VI, p. 129.
- Svestka, Z. (1976b): *Solar Flares*, D. Reidel Publishing Co.
- Tanaka, K. (1979): Prediction of great flare activity based on characteristic magnetic configurations and evolution of sunspots. P-STPW.
- Zirin, H. (1970): Active Regions. I: The occurrence of solar flares and the development of active regions. *Solar Phys.*, 14:328.

# THE NEEDS FOR PREDICTIONS AND REAL-TIME MONITORING FOR THE FLARE BUILD-UP STUDY

Zdeněk Švestka  
Space Research Laboratory  
The Astronomical Institute at Utrecht  
Beneluxlaan 21  
3527 HS UTRECHT  
The Netherlands

The paper describes the scientific aims of the Flare Build-up Study project, the organization of FBS-related solar observations during the FBS-ALERT period in May and June 1980, and the needs for predictions of solar activity during that period.

## 1. INTRODUCTION

The Flare Build-up Study is a long-term project of SCOSTEP (Scientific Committee on Solar-Terrestrial Physics) aimed at possible similarities between plasma instabilities occurring in the magnetospheric tail and in active regions on the Sun (Švestka, 1976). Intense observations of the flare build-up processes on the Sun are planned for May and June 1980 as a part of the Solar Maximum Year (SMY). Details about the organization of this active solar FBS period have been described by P. Simon and Z. Švestka in three FBS Circular Letters, copies of which can be obtained from P. Simon (Paris Observatory, Meudon) upon request. Further details can be found in SMY Newsletters published by NOAA Boulder since the beginning of 1979.

## 2. THE SCIENTIFIC PROBLEMS INVOLVED

A summary of the scientific problems to be attacked during the FBS active solar period in 1980 has been published in Appendix D of the FBS Circular Letter No. 2; relevant discussions of the science involved can be found, e.g., in Švestka (1976), Martin (1979), and Van Hoven (1978, 1979). Let us summarize here very briefly the most crucial scientific questions the FBS may try to answer:

(1) Nowadays the general opinion is that flares take their energy from the magnetic field, by simplifying the magnetic configuration in the upper atmospheric layers. Therefore, the flare build-up is understood as a build-up (i.e. accumulation and storage) of magnetic energy. However, so far no definite observational proof has been presented that this interpretation is the correct one. Thus one of

the basic tasks of the FBS is to furnish this proof.

There are two ways in which this could be achieved:

(a) First, one could try to prove that there are progressively growing shears or twists in the magnetic field that relax as a consequence of the flare process.

(b) Second, one could try to prove that current sheets are built-up in active regions and reconnect during flares.

(2) It has been well-known for almost four decades (Giovannelli, 1939) that flares occur most frequently in active regions that rapidly develop. In detail, this is manifested by a relation of flares to evolving magnetic features (Martres et al., 1968; Rust, 1972) and emerging flux regions (Vorpahl, 1973; Rust et al., 1975). Therefore, one should study carefully with high time and space resolution the relation of flares to the newly emerging flux.

(3) It is not yet clear, whether all flares represent an atmospheric response to the same kind of pre-flare situation and plasma instability, or whether there are two, or even more, different chromospheric and coronal phenomena that we include all under the common name of 'flares'.

Three examples:

(a) The frequency and importance of flares is higher in active regions with a complex (not simply bipolar) magnetic field. In particular, the occurrence of non-thermal acceleration processes greatly increases in magnetically complex active regions (cf. Švestka, 1976a). Still, flares also occur in simple (even ephemeral) bipolar regions.

(b) The existence of an impulsive burst reveals acceleration of electrons to energies of the order of hundreds of keV. In some flares these pre-accelerated electrons (and obviously also protons) are subject to a second-step acceleration process which manifests itself by radio bursts of types II and IV, and ejects relativistic electrons and high-energy protons into interplanetary space.

(c) The ATM Workshop on Flares (Moore, 1979) has indicated that there may exist two basically different types of flares: The first type, called a 'compact flare', is seen in EUV and X-rays as a static brightening of low-lying pre-existing coronal loops. The other type is the 'two-ribbon flare' that is initiated by a filament disruption and is characterized by continuous growth of brightened coronal flare loops and by progressive separation of the chromospheric ribbons in which the growing loops are rooted.

It may be that the differences between flares of different types can be verified in the differences of their build-up.

(4) Different active regions show striking differences in the productivity of flares. There are complex regions that are capable of producing hundreds of subflares, but not a single one major flare (cf. e.g.,



Švestka, 1975). Contrary to that, other regions produce not only one, but a whole series of major proton flares. Some regions produce series of tens of type III radio bursts, with and without flares, but often stop this activity after few tens of hours, without any detectable change in their magnetic configuration. Other regions produce homologous flares thus demonstrating that in some cases the magnetic configuration must be completely restored after the flaring. More generally, there are often series of flares, located in about the same position, which resemble each other in some important characteristic (e.g., microwave burst spectrum, cf. Švestka, 1975, and Castelli et al., 1974). What are the differences in the build-up of these various situations?

(5) All two-ribbon flares appear to be preceded by an activation of a quiescent filament that extends along the zero line of the longitudinal field in the active region. Of all the flare precursors, the filament activation is the only one which is clearly and uniquely associated with flares: If a dark quiescent filament becomes activated in an active region, the above-lying corona is heated and a flare-like brightening appears in the chromosphere along the filament channel within minutes to tens of minutes after the onset of the activation. In fully developed active regions the brightening is a two-ribbon flare; major events of this type are powerful accelerators of high-energy nuclei and relativistic electrons. In spotless active regions the brightening is a less energetic two-ribbon flare. In very old regions the chromospheric brightening need not be classified as a flare and eventually can be completely missing (Howard and Švestka, 1977), but it still can produce energetic coronal transients (Munro et al., 1978) and slow-mode waves (Rust and Švestka, 1979). Thus the filament activation seems to be a process of basic importance for our understanding of flares and solar activity in general, and we should study in detail the build-up of the metastable filament configurations and the agents that trigger the activations, even in cases when no chromospheric flare follows the filament disruption.

(6) Various authors have mentioned many other precursors of flares. All of them, unfortunately, can also occur without flares and, of course, most flares occur without them. Therefore, many more observations of these 'precursors' are needed, in order to be able to distinguish those of them that typically precede a flare from others that represent only fictitious, or very rare, associations. (Many of the real precursors (X-ray preflare brightenings, preheatings, EUV-line broadenings, green-line arch expansion) are probably responses of the solar corona and transition layer to the process of preflare filament activation and disruption mentioned in item (5)).

For a discussion of all these precursors and further references the reader is referred to Švestka and Simon (1969), the FBS Circular No. 2, Appendix D, and Martin (1979).

### 3. FLARE BUILD-UP INDICATORS

Many long-term precursors of flares have been mentioned in literature (see, e.g., summaries in Švestka and Simon, 1969, and Martin, 1979, and references therein). Those of them which are most reliable can be mostly seen only prior to major ('proton') flares, being various consequences of one common fact: increasing magnetic complexity in the active region that is capable to produce such flares.

Examples:

- (a) Hardening of X-ray and microwave flux. Generally, enhanced brightness of the active region at all wavelengths.
- (b) Evolution of reversely-oriented polarity configurations.
- (c) Steepening gradients at the zero line of the longitudinal magnetic field.
- (d) Increased number of flares with impulsive bursts.

One more long-term precursor has been announced by Kobrin et al. (1978), who detected an increase in radio pulsations with periods of 30-120 minutes prior to major (proton) flares. If this effect (seen for seven events) were real, it would manifest a long-lasting pre-flare instability in coronal structures. The fact that this phenomenon was seen explicitly prior to proton flares indicates that it is related to two-ribbon flares; i.e., to situations involving pre-flare filaments and their activation. Martin (1979) reports a long-lasting enhanced turbulence in filaments prior to flares which might be somehow related to these radio fluctuations.

The most reliable long-term indicator of generally high flare activity (possibly without any major flare) is the occurrence of new 'emerging flux regions', in particular if they lead to reverse magnetic polarity structures. The 'Evolving magnetic features' of Martres et al. (1968), closely associated with flare occurrences, are obviously a consequence of newly emerged flux and its interference with the preexisting magnetic configuration. Velocity field patterns indicate that flares originate near locations where the vertical velocities are zero (Harvey and Harvey, 1976), and thus flares can be expected to occur where the zero lines  $H_{||} = 0$  and  $V_{||} = 0$  intersect (Martres et al., 1971).

### 4. NEEDS FOR PREDICTIONS

Long-term precursors of this kind have been, and can be used for predictions of the flare occurrence. For the purposes of the FBS, we are interested in two kinds of forecasts:

- (a) During the two-month FBS-ALERT period (May and June, 1980) we want to know, in particular, which active regions are capable of very high flare productivity, which of them are potential candidates for the pro-



duction of a major flare, and which regions are expected to stay essentially flare inactive. Once a recommended region is selected for an FBS study, all observing stations and experiments will stay with it for several days of continuous surveillance so that it is not essential for us to know when a flare appears.

(b) At all other times, we are interested in forecasts of the time of occurrence of major flares so that intense worldwide and space observations can be initiated prior to the expected flare appearance.

There is nothing special at the forecast of type (b), because such forecasts have been made before, more or less on a regular basis. However, there are some special demands concerning the predictions of type (a) and, therefore, to make them clear, let me describe the planned activities during the FBS-ALERT period in more detail:

Starting on 1st May 1980, all participating experiments, on the ground as well as those in space will be prepared to focus all their efforts on detailed observations of a particular active region as soon as such a region is selected for the FBS study. In order to select such a region a Coordinating Center will be established at Meudon in France, with Dr. Paul Simon as the Coordinator. During the ALERT period this Center will be in daily contact with the SMM experimenters at Goddard Space Flight Center in Greenbelt, Maryland, with NOAA Boulder, Colorado and with several ground-based solar observatories which will advise the Center on the current situation on the Sun and the expected developments, and make suggestions for active regions to be selected as subjects of the FBS study.

Once a particular region is selected, it will be immediately announced in the form of an FBS-ACTION by telex (or by other fastest means wherever telex is not available) to all participating experimenters. Since that time all efforts should focus on observations of the selected region; on the ground, as well as aboard the SMM satellite all participating experiments will stay (as much as possible) all the time pointed at that region; even in case that there is a flare or another interesting phenomenon in a different region, the equipments will stay pointed on the active region selected for the FBS program.

The FBS-ACTION will continue as long as the region will be worth this intense observational effort. The shortest expected period is 2 or 3 days, the longest may be up to 10 days. After the end of the ACTION interval the ALERT period further continues and another FBS-ACTION for a different region will be announced later. Altogether we expect four to six ACTION intervals during the ALERT period which will continue through May and June.

In order to be able to deal with all the scientific problems mentioned in Section 2, we need observations of a set of active regions of various characteristic properties. Therefore, during the FBS-ALERT period, ACTIONS will be announced not only for observations of regions prolific in flares, but also for observations of other regions that may be important for the FBS program from a different point of view.



In particular, we need observations also of regions with very low flare activity (ideally without flares), to be able to decide which situations in active regions are exclusively linked to flares. Therefore, we are proposing the following (ideal) scheme of programs:

- Program 1: Study of an active region producing major flares (i.e. two-ribbon flares with fully developed loop-prominence systems, preceded by filament activation and disruption.)
- Program 2: Study of an active region very productive in small flares and subflares (but without major flares of the type described in Program 1; see example in Švestka, 1975.)
- Program 3: Study of an old spotless active region that produced major flares in the earlier rotation(s). (As a rule, such a region produces few flare brightenings, but one of them may be a typical 'spotless' major two-ribbon flare.)
- Program 4: Study of an active region inactive in flares. (For understandable reasons mentioned above.)
- Program 5: Study of a quiescent filament (prominence) which is expected to erupt. (In order to compare build-up processes associated with disrupted filaments that do, or do not, produce bright chromospheric flares.)

Programs 1, 3 and 5 would change into intense flare observations if a major flare occurs, and these observations may continue for tens of hours, in particular in space, because we are interested in the whole process of energy relaxation as the flare decays. On the other hand, during the FBS period we are not interested in studies of small flares, because we don't want to be lead away from continuous observations of the structure variations throughout the whole active region. Therefore, observations will not be changed to flare mode in programs 2 and 4 even if a flare appears; and in programs 1 and 3 the flare mode will be canceled as soon as we recognize that we are not observing a major two-ribbon flare.

## 5. THE FORECASTS REQUIRED

It is necessary to keep in mind that some participating high-resolution ground-based and sophisticated space experiments have a greatly restricted field of view, in some cases as small as  $3' \times 3'$ . Therefore, in complex active regions, the forecasts must specify in which part of the region the predicted flare activity is to be expected.

Generally, the requirements on forecasts during the FBS-ALERT period can be summarized as follows:

- (1) The earliest possible information about an active region configuration that indicates increased probability of occurrence of a major two-ribbon flare.
- (2) Regular updating of forecast (1), also with respect to the position of the expected flare.
- (3) Real-time information communicated to satellite operation centers in case that a flare appears in the region under surveillance, but it is not the major flare at which the study is aimed.
- (4) Real-time announcement of the occurrence of a major flare in the region under surveillance, to the basic network of ground-based observers and space experimenters.
- (5) Real-time updating of messages about the flare, using real-time space data for information about the post-flare loop system.
- (6) Cancellation of forecast (1) as soon as no more major flares are expected to occur in the region.
- (7) Information about expected transit of any surviving or newly developed region of type (1) over the western solar limb.
- (8) Information about reappearance of such a region from behind the eastern solar limb.
- (9) Regular updating of information (8) if the returning region is still capable of production of major flares.
- (10) Special updating of information (8) if the returning region is decayed (spotless), but still capable to produce a spotless two-ribbon flare.
- (11) If a flare appears in the recurrent region, same real-time information as in items (3), (4), and (5).
- (12) Information on any extensive quiescent filament expected to go over the eastern or western solar limb.
- (13) Updating of information (12) if the quiescent filament, during its transit across the solar disk, is suspected from becoming unstable. (Being, e.g., on the boundary of a coronal hole, cf. Webb et al. (1978); or, a new active region is being formed close to it, cf. Bruzek (1972).)
- (14) Real-time information communicated to satellite operation centers if a filament begins to be activated.
- (15) Information about an active region configuration that indicates high frequency of occurrence of subflares, with low probability of major flares. The probable location of the flaring must be given.
- (16) Updating of forecast (15), in particular if the flare activity

migrates (or is expected to migrate) to another part of the (possibly quite extensive) active region.

(17) Updating of forecast (15), and possibly its change into (1), if the active region structure indicates an increased probability of a major flare occurrence.

(18) Information about any other active region that becomes interesting from some point of view (e.g., as a source of frequent radio type III bursts).

(19) Information about all active regions that are unproductive of flares, and are expected to stay so for the coming few days.

(20) Updating of information (19), and possibly its change into (15) or (1) if the active region configuration changes.

Note added during the proof: Because of delayed launch of the SMM, there will be two more FBS alert periods in 1980, from 1 through 15 September, and from 16 through 31 October. For these two periods, the same rules are valid, as described for the May/June FBS alert interval.

#### REFERENCES

- Bruzek, A. (1952): Z. Astrophys., 31:99
- Castelli, J.P., D.A. Guidice, D.J. Forrest, and R.R. Babcock (1974): J. Geophys. Res., 79:889.
- Giovanelli, R.G. (1939): Astrophys. J., 89:555.
- Harvey, K.L., and J.W. Harvey (1976): Solar Phys., 47:233.
- Howard, R., and Z. Švestka (1977): Solar Phys., 54:65.
- Kobrin, M.M., A.I. Korshunov, S.I. Arbuzov, V.V. Pakhomov, V.M. Fridman, and Y.V. Tikhomirov (1978): Solar Phys., in press.
- Martin, S. (1979): Solar Phys., in press.
- Martres, M.J., R. Michard, I. Soru-Iscovici, and T. Tsap (1968): Solar Phys., 5:187.
- Martres, M.J., I. Soru-Escout, and J. Rayrole (1971): IAU Symposium, 43:435.
- Moore, R.L. (1979): in P.A. Sturrock (ed.), Proceedings of the Second ATM Workshop, on Solar Flares, in press.
- Munro, R.H., J.T. Gosling, E. Hildner, R.M. MacQueen, A.I. Poland, and C.L. Ross (1978): Solar Phys., 61, 201.



- Rust, D.M. (1972): Solar Phys., 25:141.
- Rust, D.M., and Z. Švestka (1979): Solar Phys., 63, 279.
- Rust, D.M., and D.F. Webb (1977): Solar Phys., 54:403.
- Rust, D.M., Y. Nakagawa, and W.M. Neupert (1975): Solar Phys., 41:397.
- Švestka, Z. (1975): ACFRL-TR-75-0437, Special Reports No. 193, p. 9.
- Švestka, Z. (1976a): Solar Flares, D. Reidel Publ. Co., Dordrecht and Boston.
- Švestka, Z. (editor) (1976b): Proceedings of the Flare Build-up Study Workshop, Solar Phys., 47:4.
- Švestka, Z. and P. Simon (1969): Solar Phys., 10:3.
- Van Hoven, G. (1978): Invited Paper presented to the Solar Physics Division of the AAS, Ann Arbor, November 1978, UCI Physics Department Report No. 78-71.
- Van Hoven, G. (1979): in P.A. Sturrock (editor), Proceedings of the Second ATM Workshop, on Solar Flares, in press.
- Vorpahl, J.A. (1973): Solar Phys., 28:115.
- Webb, D.F., P.S. McIntosh, J.T. Nolte, and C.V. Solodyna (1978): Solar Phys., 58:389.

# SOLAR FORECAST AND REAL-TIME MONITORING NEEDS OF THE STUDY OF ENERGY RELEASE IN FLARES (SERF)

David M. Rust  
American Science and Engineering, Inc.  
955 Massachusetts Avenue  
Cambridge, Massachusetts 02139 USA

This paper describes the program of the Study of Energy Release in Flares to be followed during the Solar Maximum Year. The purpose of SERF is to obtain complementary, simultaneous observations of flares from as many observatories, both ground-based and orbiting, as possible. The need for forecasts of solar activity on long-term, one-week, and two-day intervals is described. Real-time reporting is not needed, but daily summaries of activity and permanent records are important to the success of SERF.

## 1. INTRODUCTION TO SERF

The Study of Energy Release in Flares (SERF) is one of three efforts which form the scientific program of the Solar Maximum Year (SMY), which will start in August, 1979, and will run for 19 months until the end of February, 1981. The other efforts of the SMY are the Flare Build-up Study (FBS) and the Study of Traveling Interplanetary Phenomena (STIP).

The SMY has been organized to increase the effectiveness of observational research on solar activity. Many solar physicists feel that the greatest progress in solar activity research will be made with complementary, simultaneous observations in many wavebands. We feel that the needed observations can be obtained by international campaigns and agreements -- agreements aimed not simply at increasing the time or resources spent on solar research, but at coordinating observing schedules and targets according to scientific objectives established through such international forums as the International Astronomical Union (IAU) and the Scientific Committee on Solar-Terrestrial Research (SCOSTEP). In fact, organizationally, the SMY has become a SCOSTEP activity, combining the already-established STIP and FBS with the newly created SERF.

The objective of SERF is to obtain coordinated flare observations and to test current theories of how energy is released in flares. To help SERF participants work toward these goals, we intend to organize observational programs on specific scientific problems and to institute short ACTION intervals as special opportunities for observations arise throughout the SMY. The scientific study areas included within SERF are:

1. Mechanisms of thermalization of energy
2. Mechanisms of particle acceleration
3. The origins and effects of mass motions
4. Decay phase mechanisms

Some aspects of these problems were discussed in the report of the IAU Commission 10 Working Group on the SMY. These included: particle acceleration during the impulsive phase and the site of particle impact; mass motion in the impact region; second step acceleration; ejected plasmoids and their relationship to acceleration of particles; post-flare loops; and, anomalous helium ratios.

Possible mechanisms of thermalization were discussed at a SERF meeting in Glasgow in August, 1978. The mechanisms considered there were:

1. Collisions in a non-thermal beam
2. Reverse current dissipation
3. Ohmic dissipation - field annihilation
4. Shocks
5. Thermal conduction

In Glasgow, we also thought it profitable to consider subphotospheric energy sources and XUV irradiation of the chromosphere and photosphere. Descriptions of these specific mechanisms are being prepared with suggestions for observational tests. Another workshop will develop the theoretical framework and observational objectives for the other SERF problems.

### 1.1 Test Period

An "ALERT" period has been called for SERF for the latter half of May, 1979. During the ALERT, one or more ACTION intervals of 1-3 days each will be declared. During these intervals, participants will undertake intensive observations designed especially to test thermalization theories. Besides offering a test of our coordination procedures, this effort should yield useful flare data. Also, any flare data obtained during the FBS ALERT (May 1-15) can be put to good use.



Of course, the predictions of flare activity and active-region development from Boulder and Meudon will be used to help determine when and where on the solar disk observations will be made. The successes and failures of this trial run will be discussed at the IAU General Assembly in August, 1979. We expect that there will also be discussion about prediction methods, communications, and designation of ACTION initiators.

## 1.2 Operations during the SMY

The first SERF runs of the SMY will be scheduled shortly after the launch of the SMM satellite. At present, we contemplate an ALERT period for November 1-22, 1979, for all components of the SMY. The reason for this is to obtain some simultaneous ground and satellite data as early as possible in the SMM flight. Thus, should any satellite experiments fail in the first month of operation, at least the best advantage will have been taken of their period of successful operation. To this end, the SERF program will follow the SMM operations rather closely during this initial interval.

During the rest of the SMY, it appears that the following modes of collaboration will be most effective:

1. "À la carte" - observing efforts built around particular scientific objectives and particular observing opportunities. This is basically to say that the whole of the SMM is an ALERT interval for SERF (except for May and June, 1980, when the Flare Build-up Study will be the main subject of the SMY). Observing efforts (ACTION intervals) of 1-3 days will be triggered by good prospects for solar activity combined with good prospects for collaborative observations. When, for example, a large active region is predicted to give major flares and the weather is good at some major observatories and users of one or more major time-shared observatories (e.g., Solar Maximum Mission, Very Large Array, Sacramento Peak) desire to observe this activity, an ACTION interval will be called by the SERF coordinator. Although notice of the ACTION will be distributed by telex and by telephone to all interested participants, it is understood that not everyone will be able to react to every call for observations. We hope, in the course of a year's work, however, to obtain simultaneous and relatively complete coverage of several dozen flares for most, if not all, of the scientific objectives of the SERF program. ACTION intervals may be called at the average rate of 2 or 3 per month. Advance notice for ACTION intervals will vary from 1 to 7 days.

2. "Menu conseillé" - one or more pre-established intervals for especially intense efforts. This is the same as an ALERT-ACTION interval in the FBS program, and we hope that all participants will schedule their telescopes, maintenance programs, vacations, etc., so as to insure that as complete a

set of observations as possible will be obtained during this period. The candidate time for the SERF ALERT is October, 1980. During this month, the ACTION intervals should yield especially complete coverage.

### 1.3 Newsletter and Communications

World Data Center A in Boulder is preparing to publish a newsletter after the style of the present IMS newsletter to carry news of SMY activities. The first issue will appear in January, 1979. The newsletter will be mailed monthly to SMY participants according to the mailing lists of FBS, SERF, and STIP. Through the newsletter, it should be possible to broadcast and maintain a record of SMY ACTION intervals with considerable information about the observations obtained. The newsletter is composed with a computerized text editor, and all material is stored on computer files. By the use of a system of key words and dates, all references to observations of a given flare or for a given scientific objective can be recalled and summarized. A suitable code for the scientific programs, kinds of observations and observatory names will be developed from codes already in use at WDC-A and at the SMM Investigator Working Group meetings.

Volatile communications, such as telephone and telex, will be used heavily before, during and after ACTION intervals. Before an ACTION interval, participants will telex or call the SERF coordination center about observing conditions, predictions of activity, activity in progress, and recommended ACTION intervals. Before an ACTION is called, the coordinator will consult with participating observatories, according to the scientific objective of the contemplated ACTION, and especially with the following centers:

SibIZMIR (via the Sun-Earth Council)  
NOAA and the Solar Observing Optical Network  
Solar Maximum Mission Operations  
Meudon Observatory

We emphasize, however, that SERF ACTION intervals in the "à la carte" mode can be built around requests for assistance from any source, including experimenters on satellites such as ISEE-3 or Interkosmos, theoreticians who desire observational tests of hypotheses about flare energy release, and flare forecast services. For practical reasons, SERF will not normally be able to help with collaborations on problems not concerned with solar flares.

After ACTION intervals, participating observatories will be asked for a summary of observations obtained. This can be written in condensed formats, to be developed, which will describe the instrument, the observing mode, and the intervals successfully covered during the ACTION. Codes for this purpose will be included in the catalog of participants' equipment, interests,



and research programs to be issued by the Meudon Observatory for FBS and SERF. Most prospective participants have already responded to questionnaires circulated by Švestka, Zirin or Rust.

#### 1.4 Data Exchange, Collaborative Data Analysis

In contrast to the FBS plan and because of the anticipated volume of data, SERF presently does not plan to be directly involved in data analysis or data storage except to indicate to the participants which data sets seem applicable to the study of energy release in flares. Periodically, throughout the SMY, the SERF Organizing Committee and the SMY Steering Committee will review (apparent) progress made in the above-mentioned four problem areas. Periods of successful observations will be identified. Through personal contacts and via the newsletter, information on the flares most successfully observed during the ACTION intervals will be publicized. We hope to facilitate data exchanges also through meetings of international bodies such as COSPAR and SCOSTEP. We hope summaries of data obtained for outstandingly successful ACTIONS will be published by WDC-A under their series of UAG publications. Then, workshops will be organized around the four scientific objectives. The participants in the workshops will establish the level of collaborative data analysis. It is presumed, however, that all participants in the SMY will be willing to share data obtained during ACTION intervals with others via collaborative analyses and scientist-to-scientist or institution-to-institution exchanges.

## 2. NEEDS FOR FORECASTS

Because the objectives of SERF have to do with specific mechanisms of solar flares, the forecasts we need must be as detailed as possible, giving the kind of activity expected, and where and why it is expected. Most of the instruments used by SERF participants are capable of recording only one active region or segment of the solar limb at a time. Hence, predictions must be made for each active region on the disk. The "customers" are experienced solar observers with strong opinions of their own about why flares occur, and so any prediction, if it is to be used, must be accompanied by a description of recent developments on the sun, including activity summaries covering the previous 7 days.

Three kinds of forecasts are needed:



1. Active region returns (east limb)
2. Active region evolution
3. Location and type of activity over the next 48 hours

In addition, we might dream of long-term predictions, because many of the SERF facilities must be reserved months in advance of their use. I am not aware of any efforts to forecast activity several months to one year in advance, but predictions with this range would be welcomed. For example, we are now planning an intensive ALERT interval for October, 1980, but we are willing to reconsider this as late as August, 1979, during the IAU General Assembly. A prediction made at that time for activity during 1980 would be very helpful.

Outside the fixed, intensive ALERT intervals, we plan to operate in the "à la carte" mode, for which forecasts of activity covering the next 48 hours to the next two weeks will be required. First, a continually updated chart of active region returns would be very useful. It might take the form shown in Figure 1. The chart indicates at a glance how active a region was on the last disk passage, when and where it will appear on the east limb, and whether the region was growing or shrinking when last seen. This information should be useful in setting up likely ACTION intervals enough in advance to influence some telescope schedules. ACTION intervals for SERF should average about 2 - 3 per month - a full schedule for many observatories, where staffs are limited, there are competing demands for telescope time, and maintenance must be worked into the schedule.

Continuing from long term to short term forecasts, we come to active region evolution. This is probably the most important kind of prediction for SERF. We will have to establish most ACTION intervals on the basis of what we think will happen during the following 2 to 7 days. Of course, the readiness of the participating observatories enters into it, but there is no point in declaring an ACTION if nothing is expected soon. Consequently, improvements in the 2 - 7 day forecast would be most helpful, and I suspect that further research along the lines of that by Martres et al. (1968) would be fruitful. They studied the probability that certain sunspot configurations would die or become more complex and more active. As there are yet no theories for the development of active regions, further studies would have to be statistical and morphological also, but the Martres et al. work showed that some progress in predictions can be made in this way (Simon and McIntosh, 1972).

The 2 - 7 day forecasts we need must indicate what is expected from specific active regions on the disk and from those returning to the east limb. That is, the forecasts must be active region specific just as are present daily forecasts covering the approaching 48-hour period. This is because SMY participants will be interested not only in flares and bursts, but also in active region evolution and propagation of effects through slowly-changing structures

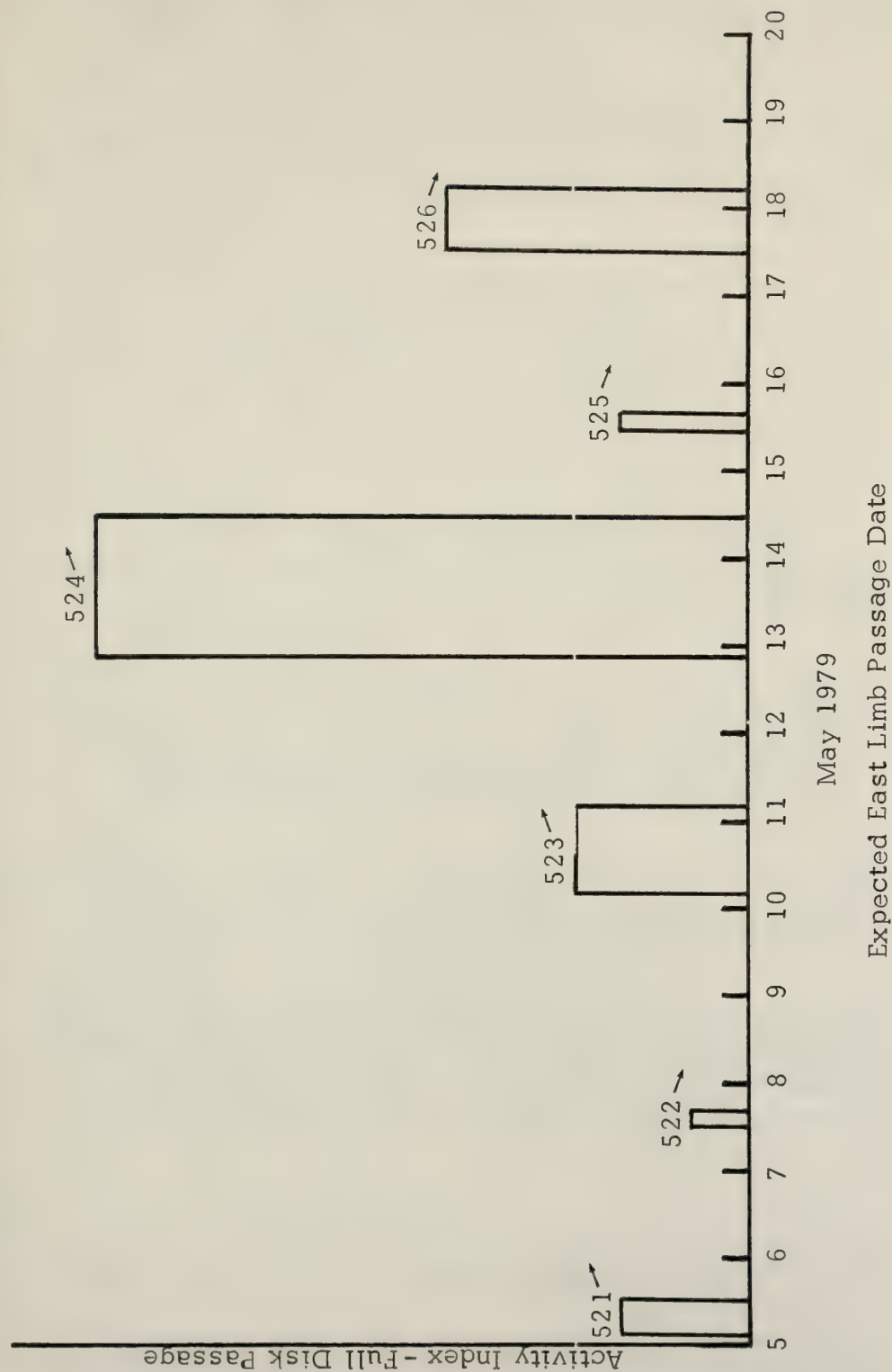


Figure 1. East-limb returns of active regions. E-W extent of each region is indicated by the width of the bars. The height is proportional to the integrated activity index for each region on its last disk passage. Numbers are McMath plage numbers. Arrows indicate growth or decay at west limb passage.

in interplanetary space. Also, 24 - 48 hour forecasts for SERF must be active region specific, because most of the research to be done during the SMY will be accomplished with instruments that can record only one active region at a time. In addition, it is important to recognize that many whole-disk instruments (e.g., some hard X-ray detectors and radio telescopes) have adjustable gain controls or recording intervals. The observer adjusts these parameters according to his science objectives, which are frequently region-specific.

Since there will be many science objectives in the study of energy release in flares, we will need appropriately detailed forecasts. Examples are regions likely to produce type III metric radio bursts, points on the limb where coronal transients are expected, filaments likely to erupt, and, of course, regions likely to produce flares. Studies by Smith (1972) and by Rust (1973) showed that it is possible to predict flaring locations to within a few heliographic degrees. That is, given that an active region is expected to produce flares, or type II bursts for that matter, it is possible to say with considerable certainty where these events will occur. This is just what is needed during the Solar Maximum Year for daily use during ALERT intervals. An ACTION interval can be called to pursue a specific study when such a specific forecast is in hand.

Solar researchers are sometimes critical of flare forecasts. Few statistical studies of flare forecast reliability have been performed, so when the forecast conflicts with a researcher's personal guess, it is difficult to gain his cooperation in a worldwide observing effort established on the basis of the forecast. So, it is important to the success of SERF and to the success of the other studies of the SMY, that the reliability of the various kinds of solar activity forecasts be documented. The present plan for selecting SERF ACTION intervals presumes agreement on the part of all participants to study the same region as the ACTION initiator. But, the manner of selecting which active region to study has been a matter of great concern, and at present we plan to use the predictions of the centers listed earlier in this paper plus the participants' opinions. However, if rational evaluations of prediction skills become available, we will surely be able to carry out our programs more successfully, by gaining wider acceptance of the targets selected for joint observations.

### 3. REAL-TIME MONITORING

If it were possible to inform all participants of flare occurrence in real time, then our observing programs could be refined considerably. The instruments aboard the Solar Maximum Mission will be programmed to operate in special "flare" modes when the hard X-ray signal exceeds a preset level.



This X-ray flare alarm can be calibrated against the standard X-ray flare classification scheme now in use at Boulder. Thus, if the solar X-ray level, available in real time at Boulder, could also be made available at SERF participants' observatories, all of us could switch to "flare" modes at the same time. This may not be as unachievable as it sounds if we accept a telex message as the indicator that a certain X-ray emission level (or other flare indicator) has been passed. I understand that the solar optical observing network is already capable of telexing a flare report within 30 seconds of flare onset.

Our greatest need in solar monitoring is not for real time reports, but for summaries of activity over the past seven days and for reliable data stored for future analysis.

#### REFERENCES

- Martres, M.-J., Michard, R., Soru-Iscovisci, I., and Tsap, T. (1968): Étude de la localisation des éruptions dans la structure magnétique évolutive des régions actives solaires. Solar Phys. 5:187.
- Rust, D. M. (1973): Estimating the flare production potential of solar active regions from analysis of real-time magnetic field data. Air Force Cambridge Res. Lab. Env. Res. Paper No. 440, L.G. Hanscom Field, Bedford, MA, USA.
- Simon, P. and McIntosh, P.S. (1972): Survey of current solar forecast centers. In Solar Activity Observations and Predictions (P.S. McIntosh and M. Dryer, eds.) MIT Press, Cambridge, USA, 343 - 358.
- Smith, J.B., Jr. (1972): Predicting activity levels for specific locations within solar active regions. In Solar Activity Observations and Predictions (P.S. McIntosh and M. Dryer, eds.) MIT Press, Cambridge, USA, 429 - 442.

REQUIREMENTS FOR PREDICTIONS AND REAL-TIME MONITORING  
FOR THE STUDY OF TRAVELLING INTERPLANETARY PHENOMENA

Murray Dryer  
Space Environment Laboratory  
NOAA-ERL  
Boulder, Colorado 80303, USA

and

M. A. Shea  
Space Physics Laboratory  
Air Force Geophysics Laboratory  
Bedford, Massachusetts 01731, USA

This paper summarizes some of the scientific objectives of the Study of Travelling Interplanetary Phenomena particularly during the years of the solar maximum of Solar Cycle 21. These objectives are concerned with the understanding of physical processes within quiescent as well as disturbed flows of solar wind plasma, energetic particles, and magnetic and electric fields from the sun. It is noted that predictions, while needed for long-range logistical planning, is secondary to the requirements of synoptic real-time monitoring over a wide range of particle energies and spectral bandwidths.

## 1. INTRODUCTION

The Study of Travelling Interplanetary Phenomena (hereafter referred to as STIP) is one of the long-term projects of the Scientific Committee on Solar-Terrestrial Physics (SCOSTEP). Convened in 1973 and as described by Dryer and Shea (1976), STIP is directed toward the interdisciplinary research of interplanetary physics via the mode of informal international projects which often are set up on an extemporaneous basis among scientists of diverse disciplines. All of STIP's activities have been documented in a series of STIP NEWSLETTERS which may be obtained from one of the authors (M. A. Shea). Briefly, these activities begin at the sun where solar physics, insofar as that subject concerns "quiet" and "disturbed" ejecta of all kinds into the interplanetary medium, is of obvious interest. Other activities are concerned with energetic particles which range from the hundreds of keV to the tens of GeV - that is, those particles which are ejected from the sun at non-thermal energies in the sub-relativistic and even relativistic ranges. The sub-relativistic particles may be even accelerated locally by in situ electric fields. Thermal plasma (the solar wind), magnetic and electric

fields, together with the study of plasma waves are also of interest to STIP participants. The role of solar radio astronomy and interplanetary scintillations of signals from artificial and natural point sources of electromagnetic signals is an important, remote sensing counterpart of in situ spacecraft observations. Solar wind interaction with planetary magnetospheres (such as those of Mercury, Earth, Mars, Jupiter and Saturn), planetary ionospheres (such as that of Venus), and comets is also a part of STIP activities. Such interaction, particularly during disturbed periods of time, is important because of the response celestial obstacles may undergo such as sudden storm commencements, substorms, ionospheric deformations, radio emission (e.g. non-lo dekametric emission from Jupiter), and possibly albedo changes due to chemical reactions in ionospheres and atmospheres of comets (e.g. Giocobinni-Zinner), planets (e.g. Neptune), and natural satellites (e.g. Titan).

By its very nature, STIP encompasses much of the scientific thrust of solar-terrestrial and space plasma physics. The study of the sun itself (the build-up of conditions conducive to solar flares and the energy release process itself—as studied by SCOSTEP's FBS and SERF projects during the Solar Maximum Year), magnetospheric physics, ionospheric physics, and the coupling of the latter two links in this chain with the neutral terrestrial atmosphere are, of course, not included per se in STIP research.

Examples of STIP research are given in an early description of the project's philosophy (Dryer and Shea, 1976). In that paper, the authors used examples from their own immediate research (solar wind disturbances and solar cosmic rays) as examples of the wider range of topics noted above and which are naturally of research interest. The goal of any research is obviously the achievement of knowledge of various physical processes. In the case of STIP, this understanding must be verified, checked, and confirmed by many observers and theoreticians. The ultimate test of any theory, of course, concerns its ability to provide predictive accuracy. An example of this approach is outlined later in this paper.

## 2. SCIENTIFIC OBJECTIVES

The scientific problems with which STIP participants are concerned range over the spectrum of space plasma physics. Transport of mass, momentum and energy takes place throughout the solar system (c.f., Space Science Board Report, S. Colgate, Chairman, 1978) under both quiescent and disturbed conditions. The quiescent state must always be established and understood in order that spatial and temporal deviations from that state could be described and confidently predicted. The chain of events, so to speak, that takes place between the sun and the earth's magnetosphere must be closely monitored and modelled so that terrestrial perturbations can be forecast on the basis of proven physical facts.

Examples of current interest may be mentioned. If we can prove that the interplanetary Poynting flux,  $\underline{ExB}$ , is coupled into the magnetosphere such that substorm activity can be triggered, we may use the energy parameter,  $\epsilon = VB^2 \sin^4(\theta/4) \ell_o^2$ , ( $\text{erg sec}^{-1}$ ) which Perreault and Akasofu (1978) show to be correlated quite favorably with the Joule dissipation (indicated



by the AE index) in the auroral zones. That is, a monitoring spacecraft (such as the Pioneer 6-9 series or ISEE-3) could provide  $\sim 1$  hour advance notice of the solar wind velocity,  $V$ ; total magnetic field,  $B$ ; and  $\theta$  ( $= \tan^{-1} |B_y/B_z|$  for positive  $B_z$ ; and  $\theta = \pi - \tan^{-1} |B_y/B_z|$  for negative  $B_z$ ). In the above equation the  $\ell_0$  is the reference magnetospheric radius. An alternative lower latitude substorm correlation parameter is  $V^2 B_z$ , suggested by Maezawa (1978), which also is related strongly to southward fields as well as to the solar wind dynamic pressure ( $1/2 \rho V^2$ ). The monitoring philosophy noted above is relevant, of course, to both of these parameters (and even, possibly, others). Physical understanding will be enhanced by the use of MHD numerical simulation models in which the quiescent state of the sun and solar wind, followed by adequately-described perturbations in the "transmission line," are tested in direct confrontations with observations.

Another scientific problem is the description of solar energetic particle fluxes under the same kinds of conditions noted above. It is essential that the solar sources of background fluxes be identified via close collaboration of ground/space solar observatories and in situ spacecraft observations outside the earth's bow shock wave. Scientific questions concerning coronal transport of energetic particles are still being debated. Do these particles escape directly from active regions, or if prohibited from doing so, do they diffuse through large longitudinal distances before escaping via open field lines? Does the latter process also take place during and following solar flares? What are the roles of flare-caused magnetic field deformation and shock wave generation on the acceleration of particles to sub-relativistic energies? A grid of heliocentric spacecraft (such as a configuration of Helios, ISEE, Pioneer and Venera spacecraft) would be of value in testing some of the proposed models. The same grid system will be of essential necessity in order to make it possible to determine the ecliptic plane's response to a solar flare or coronal hole. Multi-site radio astronomical observatories which monitor the interplanetary scintillation patterns of EM radiation from distant pulsars and quasars are important partners in the determination of shock and piston velocities and global extent of such disturbances above and below the ecliptic plane. Ground-based radio observatories which measure spectral bursts (such as Types II, III and IV) as well as the spatial extent of fixed-frequency sources close to the sun are also indispensable and complementary partners in such synergistic observations. Space-borne radio receivers at longer wavelengths aboard spacecraft which have spin axes both perpendicular as well as within the ecliptic plane can track the electrons responsible for stimulating Type III radio emission and, in effect, plot the location of solar magnetic field lines. The latter can often follow a warped path which follows the general "ballerina-skirt" solar current sheet which, in its east-to-west turning with the sun, is detected as magnetic sector changes.

Other scientific problems abound. For example, does magnetic reconnection take place in interplanetary space? Do coronal transients always get "pinched-off" to create bubbles which expand as they move outward through space? Or, alternatively, are such examples rare events, and do the transients generally propagate as monstrous waves upon a pre-existing mildly-disturbed solar wind sea? The answer is of vital importance to forecasters for, if the latter suggestion is valid, it will be much easier to produce MHD

models which -- when perturbed at the sun -- produce the magnetic field flip-flopping (due to Lorentzian forces) responsible for changes in the Poynting energy flux discussed briefly above. An example of STIP's need for monitoring, rather than forecasting, is given below.

### 3. AN EXTEMPORANEOUS EXAMPLE OF MONITORING INTERPLANETARY CONDITIONS

The second and third issues of the Solar Maximum Year's Newsletter (SMY-NL) described the very-preliminary observations in December 1978 of several large flares, at least one of which was responsible for a shock wave detected at Venus by Pioneer-Venus and then as an SC at earth. The third SMY-NL editorialized that "this chain of events from solar activity through cooperative notification, special configuration of experiments, special processing of data and prediction of effects substantiated by measured results provides an example of one type of activity that may be particularly productive during the forthcoming Solar Maximum Year."

The chain of events was as follows. On 11 December 1978, following several days of very minor activity, an optical 2B flare at S16°W48° (McMath Region 1447) began at 1807 UT, reached a maximum at 1925 UT, and ended at 2325 UT. Its X-ray flux was classified as M7. Nearly simultaneously, an optical 2B flare (X-rays: X1) began at 1833 UT at S17°E14° in McMath Region 1444. The latter flare reached maximum at 1942 UT and ended at 2018 UT. The Culgoora radioheliograph reported to the Boulder SESC that a major stationary Type IV burst took place from 3800 MHz down to 6 MHz. Shortly afterward, Dr. J. A. Joselyn of the SESC staff notified one of the present authors (MD) of the above events. Operating on the hypothesis that the radio burst originated from the S16°W48° flare (no information on the exact timing of the radio burst starting time was available at that time), MD informed the PI (Dr. J. H. Wolfe) of the solar wind plasma experiment (NASA/Ames) on Pioneer-Venus which had been inserted into Venusian orbit about a week earlier (5 December 1978). At that time, the NASA/Ames experiment was measuring the decline of a high-speed stream. Venus at 0.7 AU was close to the central meridian of the S16°W48° flare. The quick-look solar wind data was thereupon closely monitored by Wolfe and his co-investigators around the expected arrival time of the flare's shock wave at Venus. Due to a 5-hour data gap, the exact arrival time of the shock was not known. However, the high probability that it had indeed arrived was indicated by a large step-like increase from a  $\sim 250 \text{ km sec}^{-1}$  minimum at the end of the earlier, high-speed stream to  $\sim 570 \text{ km sec}^{-1}$  at the end of the data gap on 13 December. The solar wind velocity decreased fairly quickly thereafter during the next few days indicating the passage of a blast-produced shock wave. Upon detection of this shock, Wolfe (private communication, 13 December 1978) predicted the probable arrival time of the shock at earth assuming constant shock velocity ( $\sim 850 \text{ km sec}^{-1}$  as an average, based upon the start time of the S16°W48° flare and the detection of the  $570 \text{ km sec}^{-1}$  plasma velocity at Venus) between Venus and earth. He also predicted the  $\Delta H$  of the SSC at earth of 60 nT. The success of the prediction was confirmed when the SSC was recorded at earth about 1.5 hr after the predicted time of arrival about 0200 UT on 14 December 1978. Also, the magnetic field increase recorded near local noon was 63 nT.



Pioneering efforts such as this one indicate the utility of having prompt arrival of solar wind plasma, magnetic fields, and energetic particle data (as provided in the past by the Pioneer 6-9 series) to the forecast and monitoring centers such as Boulder's SESC. As additional data associated with the 11-14 December 1978 example described briefly above become available to STIP participants, it is expected that additional illumination on scientific questions in the interplanetary medium will take place. The reader may well ask if one may unambiguously rule out the  $S17^{\circ}E14^{\circ}$  flare as the perpetrator of the shock. The answer is no because there are, as yet, no data available from ISEE-1, -2, -3 or Prognoz-7 which will help to resolve this question. Nor were there any spacecraft (to our knowledge) above the eastern solar hemisphere at the time. However, preliminary IPS observations from UCSD and Toyokawa are available (Dr. T. Watanabe, private communication, 1979). These two multi-site observatories detected a solar wind profile along the line of sight to 3C298 ( $34^{\circ}$  above the ecliptic) which mimicked, albeit delayed, the NASA/Ames experiment. The spatially-located peak of the scintillation profile was very close to the position of Pioneer-Venus. Obviously, conclusive proof of the "western flare" hypothesis is still to be forthcoming. We hope, however, that the value of extemporaneous, coordinated, informal, international and interdisciplinary cooperation espoused by the STIP philosophy will have been accepted by the reader. Also, it should be stated that synoptic monitoring is a first priority for STIP's scientific objectives. Forecasting events, such as solar flares, is indeed also of importance for altering various kinds of observatories. Forecasting energetic particle and solar wind plasma/field events at earth or at other celestial obstacles is, obviously, the ultimate goal as stated earlier. How closely the monitoring vis-à-vis forecasting goals go hand-in-hand is a process which cannot be predicted very successfully at this time. Hence our emphasis on the former, more modest goal.

#### 4. STIP OPERATIONS DURING THE SMY

We have continued our practice of declaring, on the basis of common consent at various STIP Business Meetings, STIP INTERVALS in advance. Summaries of the scientific results from STIP INTERVALS and Retrospective World Intervals are contained in the D. Reidel, AFGL and UAG proceedings of various SCOSTEP/COSPAR/STIP symposia and in Space Science Reviews (1976), the latter containing a summary of the famous August 1972 events. The interested reader may inquire from the STIP Secretary (MAS) or WDC-A as to the availability of Newsletters, etc.

A practice ALERT for SMY will occur during STIP INTERVAL VI (15 April-15 June 1979). This ALERT was chosen in part because of a favorable Helios-1 and Helios-2 configuration and also because of the simultaneity of FBS and SERF practice ALERTS as described elsewhere in this volume by Svestka (1979) and Rust (1979) with respect to solar activity per se. Results of STIP INTERVAL VI, including mistakes and potential improvements, will be outlined during a special session of IAU Commission 10 at the 1979 General Assembly.

STIP INTERVAL VII (August-September 1979) will constitute the beginning of the SMY. This period, nick-named the "Spacecraft Parade," was chosen because of the passage, from left to right, over the sun's north pole of Voyager-1, Voyager-2, and Pioneer-11 (the latter involved in a Saturnian fly-



by during this time), with Pioneer-Venus (still in orbit about Venus) passing over the pole from right to left. Also, Helios-1 and Helios-2 will both be "viewing" the sun's western hemisphere while ISEE-3 will monitor the flows toward earth. The telemetry lines-of-sight to the polar "spacecraft parade" will be inclined about  $1^{\circ}$ - $2^{\circ}$  above the sun's pole. Determination of high latitude solar wind velocities and densities are theoretically possible from the two-frequency telemetry link to these spacecraft. A STIP scientific objective -- the latitudinal extent of quiescent and disturbed flows at polar latitudes -- may possibly be achieved at this time. The cooperation of tracking and data acquisition facilities, spacecraft investigator groups, and ground-based observatories is obviously of great importance.

The SMY Steering Committee is presently investigating the likelihood of additional STIP INTERVALS during the SMY (1 August 1979-28 February 1981). Such intervals will undoubtedly be influenced by spacecraft constellations along radial and/or spiral paths. Needless to say, we can expect solar "self-declared" INTERVALS which emphasizes the importance of extemporaneous, informal, and fairly rapid cooperative efforts in communication (to and from the Boulder, SESC and Meudon, for example) and cooperation. The advantage of the advance-declared INTERVALS is, of course, achieved sometimes when national agencies find it possible to prepare their tracking schedules far in advance to accommodate a SCOSTEP/STIP request. The indispensable assistance of the NASA/Goddard NSSDC and WDC-A will be of major importance in the advance declarations -- which will be announced in the SMY-NL and STIP-NL.

## 5. CONCLUDING REMARK

The multi-disciplinary nature of STIP's scientific objectives and the vast region of space within which it operates makes it imperative to emphasize its informal operating procedure. One cannot easily organize diverse national organizations and diverse disciplines into a smoothly operating, quick-response, operation which involves multivarious forms of advanced technology. It is, instead, hoped that the cooperative monitoring efforts of these groups are essential to the common denominator of all their objectives -- the study of the many fascinating aspects of travelling interplanetary phenomena and their relevance to terrestrial disturbances.

## REFERENCES

- Dryer, M. and Shea, M. A. (1976): Solar Phys., 47:413.
- Maezawa, K. (1978): Presented at the AGU Chapman Conf. on Quantitative Modeling of Magnetospheric Processes, La Jolla, California, Sept. 1978.
- Perreault, P. and Akasofu, S.-I. (1978): Geophys. J. Ast. Soc., 54:547.
- Rust, D. M. and Švestka, Z. (1979): this volume.
- Space Science Board (S. A. Colgate, Chairman, Study Committee) (1978): Space Plasma Physics, 1, Study Committee and Advocacy Panels Rept.

V. MAGNETOSPHERIC PREDICTIONS  
A. SOLAR WIND AND MAGNETOSPHERE INTERACTIONS

A WORKING GROUP REPORT prepared by C. T. Russell, Chairman, Members: J. H. Allen, D. P. Cauffman, J. Feynman, E. W. Greenstadt, R. E. Holzer, S. M. Kaye, J. A. Slavin, R. H. Manka, G. Rostoker and W. F. Stuart.

We now have the ability to characterize the state of the magnetosphere quantitatively and to predict this state from a knowledge of interplanetary conditions. We recommend therefore that the conditions of the solar wind and interplanetary medium be continuously monitored, in addition to monitoring the state of the magnetosphere. We further recommend continued study of the relationship of interplanetary conditions to the response of the magnetosphere, continued study of the geomagnetic tail, tests of Pc 3,4 magnetic pulsations as diagnostics of the solar wind and tests of kilometric radiation as a remote monitor of the auroral electrojet. Finally we urge the continued construction of geomagnetic indices and the development of new indices where there is a proven advantage to these new indices.

1. INTRODUCTION

Within the overall Solar Terrestrial system, the interface between the solar wind and the magnetosphere is crucial since it is this interface which determines how much of the solar plasma and field energy is transferred to the earth's environment. This coupling not only varies in time, responding to major solar disturbances, but also to small changes in solar wind conditions and interplanetary field directions. We do not yet fully understand this coupling, though we have some intriguing clues, and yet it is upon this coupling that we base part of our ability to make solar terrestrial predictions. Due to the rapidly evolving nature of this science on the one hand, and the considerable unknowns and complexity on the other, further intensive study of the Solar Terrestrial system is necessary if we are to increase our ability to understand the response of the earth to the sun.

If we are to predict the response of the magnetosphere to some disturbance in the solar wind, we must first have some accurate knowledge of what that disturbance will be. The longest warning one could expect would be from solar observations, but at present our ability to predict the properties of the solar wind at one astronomical unit from observations of the sun are far from adequate for the purpose except for the largest disturbances. We hope that future studies will improve this situation and we urge that these be carried out.

Interplanetary scintillation measurements of radio sources near the sun also promise to provide some warning, but these also have limitations. They provide measurements of the solar wind velocity and turbulence but not magnetic field strength and direction. Furthermore, the radio sources necessary for such studies are not distributed uniformly in the sky. On the other hand, computer modeling studies show that measurements or inferences of solar wind velocity density and magnetic field can be extrapolated over large radial



distances. Thus, if measurements of the solar wind close to the sun were available, they could be useful for long lead-time forecasts. At present we must use solar wind and interplanetary field measurements made close to the earth for accurate predictions of magnetospheric response. As we discuss below, it is important for predictions that we better understand the distance over which solar wind conditions can be extrapolated.

If measurements outside the magnetosphere are unavailable, observations of the present state of the magnetosphere can be useful in predictions since the magnetosphere can store energy before releasing it into the ionosphere. Also geomagnetic pulsations are thought to provide diagnostic information both on the present state of the magnetosphere and to a limited extent on external conditions.

If we have an adequate measure of the solar wind conditions, we can attempt to predict magnetospheric response. This area is still under active investigation but enough is known to provide some practical information. In the past, in the absence of solar wind data, many forecasters used geomagnetic indices to predict the reaction of the ionosphere, for example, to disturbances of the magnetosphere. It now seems possible to bypass the indices for this particular purpose. However, before doing so we should examine the relationship between these indices and the solar wind parameters that control them. We do this in the third section of the report.

Finally we list a number of recommendations. Forecasting magnetospheric conditions promises to become increasingly important in the years to come as technology advances. Thus, it becomes increasingly important to monitor the conditions that affect the magnetospheric environment and to pursue studies which will provide a better understanding of this environment and an improvement in our ability to predict its state.

## 2. MEASUREMENTS AND INFERENCES OF SOLAR WIND CONDITIONS

For accurate predictions we need accurate measurements of the ambient solar wind conditions. However, even in the absence of measurements, we may still deduce some information on these conditions by examining the state of the magnetosphere itself. In this section of the report we discuss the requirements for in-situ observations and the information available from indirect techniques.

### 2.1 In-Situ Solar Wind Conditions at 1 AU

In order to predict the response of the magnetosphere at any particular time, we must know the conditions in the solar wind external to the magnetopause. In particular, it has been shown that geomagnetic activity is principally controlled by the solar wind velocity and the interplanetary magnetic field strength and direction (Burton et al., 1975; Murayama, 1979; Maezawa, 1979). Other parameters such as density may play some role but are generally thought to be less important. With the exception of the occurrence of shock waves, the velocity and magnetic field strength are slowly changing parameters



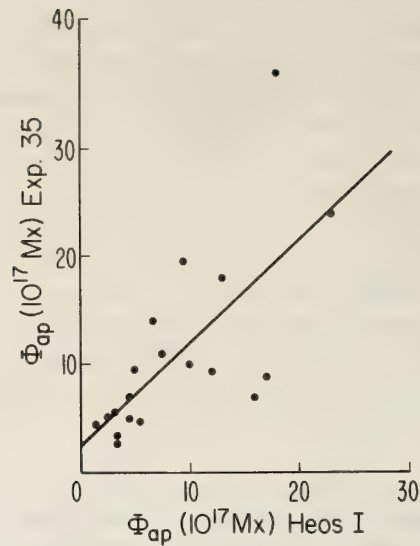


Figure 1. Comparison of interplanetary magnetic flux measured by two near-earth satellites. The correlation coefficient is 0.70. The spacecrafts were always separated by less than 100 earth radii (Holzer and Slavin, 1979).

when compared with the magnetic field direction. Thus, while the strength of the interplanetary field and the solar wind velocity may be very important parameters in controlling geomagnetic activity, it is the changes in field direction which are correlated with geomagnetic 'events' in the magnetosphere. Furthermore, since the field of the solar wind changes direction often, its correlation length or region of uniformity is relatively small, requiring that the measurements of solar wind conditions that lead to small or moderate geomagnetic activity be made relatively close to the earth.

Ideally we would like to monitor the element of solar wind plasma that is going to intercept the earth. If we do not and if the planes of discontinuity are not normal to the solar wind velocity, the discontinuities will reach the earth with variable time delays depending on the location of the satellite with respect to the sun-earth line and the orientation of the discontinuity plane. Such variable delays have been deduced for earth-orbiting satellites (Perreault and Kamide, 1978; Pytte, 1978) and the correlation coefficients between spacecraft separated by as little as 80 Re was found to be low by (Chang and Nishida, 1973). Finally as shown in Figure 1 Holzer and Slavin (1979) have made comparisons between time integrated values of  $V_{sw} B_s$ , as might be used in forecasting, observed at HEOS 1 and Explorer 35. This comparison shows a good average agreement but a surprisingly low correlation coefficient of 0.7. We feel that it is important to develop a clearer understanding of how the solar wind plasma and field conditions change along and perpendicular to the solar wind flow. The above studies also indicate that it is important to consider carefully the placement of satellites for solar wind monitoring. We leave the question of the functional form of the dependence of geomagnetic activity to a later section of the report.

The interaction region which is thought to be important for geomagnetic activity is generally believed to be the forward hemisphere of the magnetosphere. To probe directly the magnetosheath plasma in this region would require many satellites in circular  $\sim 12$  earth radii (Re) circular orbits to ensure that one was always near the interaction region. Clearly this is impractical. A more practical approach would be an  $\sim 30$  Re circular orbit like IMP's 7 and 8. However, this would require at least 2 satellites in earth orbit. Another proposal is to monitor at the forward libration point 235 Re along the earth-sun line. However, for communication purposes a satellite cannot be placed exactly here. It must orbit around the libration point. ISEE-3 does this but with an 100 Re semi-major axis; tighter orbits are possible. A further possibility is a lunar 'synchronous' orbit which maintains apogee in the solar wind using close encounters with the moon.

At present the ISEE-3 spacecraft is observing the solar wind in a large,  $\sim 100$  Re semi-major axis, ellipse centered on the forward libration point. Comparisons of the data obtained on this spacecraft with that obtained closer to the earth on ISEE 1 and 2 should be undertaken as soon as the requisite data become available to address this problem. It is expected that a high correlation coefficient will result for the largest and long enduring events, but, based on the studies mentioned above, that small events will correlate poorly.

## 2.2 Measurements Within the Magnetosphere

The earth's magnetotail maps into the polar cap. Thus, changes in the magnetotail are reflected in changes in the polar cap. Changes in the size of the auroral oval can be monitored by spacecraft TV cameras, radar measurements, satellite particle detectors, and ground based magnetometers. Ground based instrumentation suffer the problem of being point, or at best line, measurements and thus must be treated with some caution, taking into account the shape of the auroral oval. In all these measurements it is easier to distinguish the equatorward border of the auroral oval than the poleward border. However, even an approximation to the size of the polar cap should prove useful.

Figure 2 shows the response of the lower border of the auroral oval to an increasingly southward  $B_z$  (Kamide and Winningham, 1977). Burch (1972, 1973) has examined the motion of the polar cusp on the dayside in a similar way. These studies suggest that a measure of the size of the polar cap would be a very useful index of the state of the magnetosphere. However, it may not solely vary in response to the solar wind but also due to internal magnetospheric affects such as auroral activation and changes in the ring current (Siscoe, 1979).

In addition to measuring the size of the polar cap, it is useful to measure plasma flows across the polar cap. Increased stress on the magnetosphere causes flows across the polar cap from noon to midnight. These flows in turn cause currents. These plasma flows can be monitored by low altitude satellites measuring electric fields, by high altitude balloons equipped to measure electric fields and with ground based radars. Low altitude polar orbiting spacecraft spend about 20% of their orbit in the polar cap and could provide



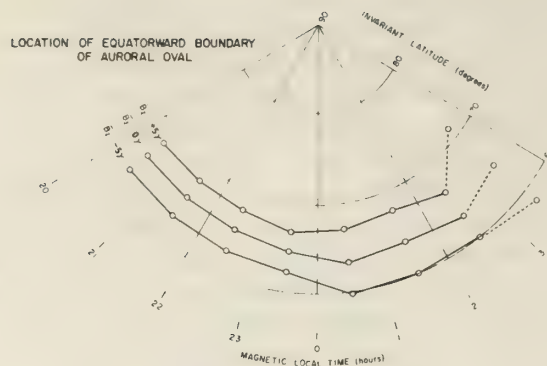


Figure 2. Location of the equatorward boundary of the diffuse electron precipitation at different local times in the dark sector for three  $B_z$  values of IMF (Kamide and Winningham, 1977).

a measure of the potential drop across one or the other polar cap about every 50 minutes. Balloons can provide continuous point measurements, but have lifetimes of the order of one day. Scientific radars presently in place, monitor at best only the edges of the polar cap. Magnetometers can measure the resultant currents, but the current systems are complicated and a function of local time. Thus a longitudinal and meridional high latitude network is necessary for 24 hour-a-day coverage.

In the outer magnetosphere increased stress leads to changes in the magnetic field (McPherron et al., 1973). When a substorm occurs, this stress is relaxed and the field configuration returns to a dipolar configuration. This presubstorm phase has been called the growth phase (McPherron, 1970). A very sensitive location in the magnetosphere for measuring these field configuration changes is at synchronous orbit. This can be done with magnetic field measurements or energetic electron measurements. Prior to a substorm, 97% of the time, energetic electrons assume a cigar-shaped distribution about the field direction (Baker et al., 1978, 1979). Thus, energetic particle measurements at synchronous orbit can be used to forecast the injection of fresh energetic particles.

### 2.3 Geomagnetic Pulsations

The predictive value of geomagnetic pulsations for space science studies or for practical applications is presently undetermined or undeveloped. Pulsations are thought to be able to provide diagnostics of conditions in the distant magnetosphere and the solar wind. Thus, their occurrence rates, periods and amplitudes have been correlated with a wide variety of magnetospheric and solar wind parameters such as the distance to the magnetopause (Vero, 1975), the location of the plasmopause (Orr, 1975), the density in the plasma trough (Orr, 1979), the location of the polar cusp (Bolskakova et al., 1975); the strength of the interplanetary magnetic field (Gul'el'mi, 1974); and its direction (Troitskaya et al., 1971), and the solar wind velocity (Singer et al., 1977). In fact, it is the sensitivity of geomagnetic pulsations to a wide variety of different parameters in the solar wind and magnetosphere, that impairs their present predictive ability, while still offering hope of future profitable applications.



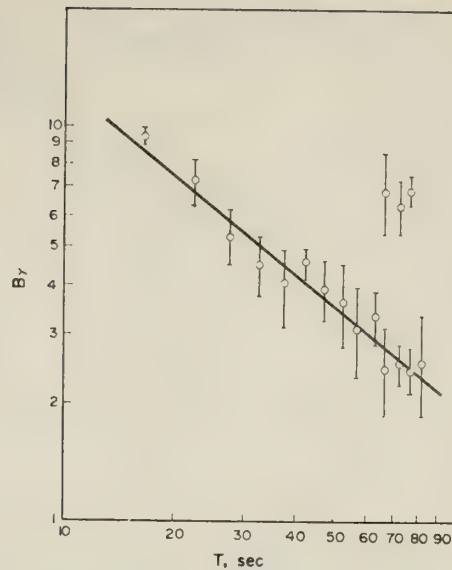


Figure 3. Average interplanetary magnetic field strength versus period of magnetic pulsations seen at a mid-latitude observatory (Troitskaya et al., 1972).

One prevailing model of the generation of these waves proposes that any perturbation at the magnetopause under suitable conditions becomes amplified into a surface wave, via the Kelvin-Helmholtz instability. The waves then couple into the magnetosphere and resonate as standing waves along field lines (Southwood, 1968). A second model suggests that the waves originate as part of the quasi-parallel structure of the shock (Greenstadt, 1972). Quasi-parallel shock associated waves are always present somewhere. When the interplanetary magnetic field is nearly aligned with the velocity of the solar wind, the unstable region moves closer to noon. The waves are convected into the subsolar magnetosheath and blown against the magnetopause. The motion of the magnetopause in response to these waves couples the energy into the magnetosphere and the wave energy builds up wherever the frequency of the disturbance equals a resonant frequency of the field line. Neither model precludes the validity of the other and indeed there is every reason to believe they both apply simultaneously (Wolfe et al., 1978; Greenstadt et al., 1979). Thus, pulsations may offer an opportunity to characterize at least some properties of the solar wind, magnetosheath and magnetopause when direct measurements are unavailable.

Figure 3 shows one of the clearest relationships, that of pulsation period versus interplanetary magnetic field strength (Troitskaya et al., 1972). This average correlation was strong enough to convince the observers at Borok Geophysical Observatory to calculate an index of interplanetary field strength. The validity of this index was later tested independently (Russell and Fleming, 1976). The results of this test are shown in Figure 4. The interplanetary field strength correlates well with the index but there is significant overlap between the distribution of field values for each level of the index. Thus although the average correlation is high, the accuracy of the index as a predictor of individual hourly field values is low.

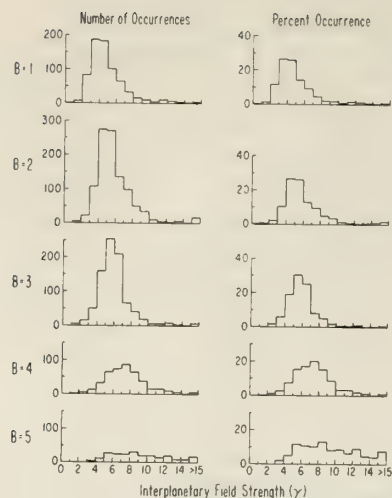


Figure 4. Histograms of the occurrence of interplanetary field strengths for each level of the Borok B index (Russell and Fleming, 1976).

The advantage of pulsation measurements is that they are relatively inexpensive to make. They can be obtained from a modest midlatitude world-wide network with perhaps as few as eight stations. However, this advantage must be weighed against the present low accuracy of the resulting inferences. It is possible that future studies, especially those of waves in space, may lead to a better understanding of the sources of magnetic pulsations, and their properties. In particular, a better knowledge of the conditions under which hydromagnetic waves are generated at the magnetopause and how they propagate through the magnetosphere would be useful. Perhaps studies of wave polarization might assist in distinguishing pulsations with differing sources. Then perhaps the diagnostic capability of magnetic pulsations can be improved. Thus we recommend the continued investigation of the relationships between interplanetary parameters and pulsation amplitudes and periods but cannot define the use of pulsations in magnetospheric forecasts at this time.

### 3. RESPONSE OF THE MAGNETOSPHERE TO SOLAR WIND CONDITIONS

We shall assume that we have a measure of the solar wind velocity and density and interplanetary magnetic field magnitude and direction close to the location of the earth. As mentioned above, in order to make a confident prediction of the geomagnetic response to these conditions, we must be sure that the scale length for variations in the critical parameters are longer than the impact "parameters" of the element of the solar wind being probed. We currently do not have an adequate measure of these lengths, but satisfactory predictions have been made from data obtained within about 50 Re of the earth in the past.

The earth's magnetosphere responds to changes in dynamic pressure of the solar wind, by changing its size. If the magnetosphere is in a stressed state, compressions may lead to a substorm. Rapid changes cause sudden impulses in ground records and are classified as a form of geomagnetic activity. However, these compressions, per se, are not usually associated with the cur-



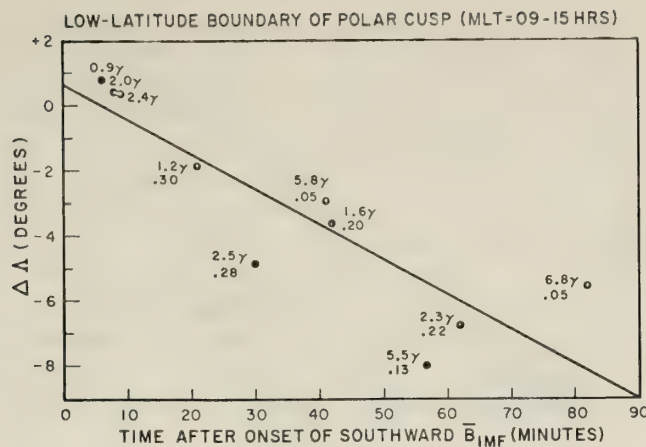


Figure 5. Equatorward motion of the polar cusp after a southward turning of the interplanetary magnetic field (Burch, 1972).

rent systems and energy deposition arising during storms and substorms. If the interplanetary field is southward at the time of an interplanetary shock passage or turns southward sometime later, then the IMF is likely to be strong due to compression by the shock, and significant geomagnetic activity in the form of a geomagnetic storm will result (Burton et al., 1975).

Pressure changes alone are not the only cause of variations in the magnetopause location. Southward turning of the IMF also causes the magnetopause to move inward (Aubry et al., 1970). Figure 5 shows this effect in the motion of the polar cusp (Burch, 1972). However, in this process the magnetosphere does not change size so much as it does shape. The dayside magnetosphere loses magnetic flux which enters the polar cap and hence resides in the magnetotail. The convection across the polar cap causes field aligned current systems to arise which close in the polar cap. Asymmetries may arise in the flow across the polar cap, or equivalently the electric field and the currents, depending on the sign and magnitude of the interplanetary eastwest component of the interplanetary magnetic field (Heppner, 1972).

At lower latitudes the electric field is also found to be controlled by the IMF. Horwitz et al. (1978) using the Chatanika radar found that convection in the F-layer from  $\lambda = 63^\circ$  to  $68^\circ$  was proportional to the  $B_z$  component of the IMF. At the highest latitudes this response was simultaneous with the IMF change but at the lowest an  $\approx 30$  minute delay was observed. Blanc (1978) has studied the field penetration to even lower latitudes with the Saint-Santin radar ( $\lambda = 47^\circ$ ). Quiet time east-west drift velocities of increased from 10 m/s to  $\sim 100$  m/sec within 30 min after  $B_z$  turned southward. Estimates of the convected flux returned to the dayside correlate quite well with the applied potential or mergible flux in the solar wind as shown in Figure 6 (Holzer and Slavin, 1979). These measurements are consistent with the studies of injection into the partial ring current (Clauer and McPherron, 1979) and the ring current (Burton et al., 1975) who showed that injection was proportional to the southward  $B_z$  and the solar wind velocity.



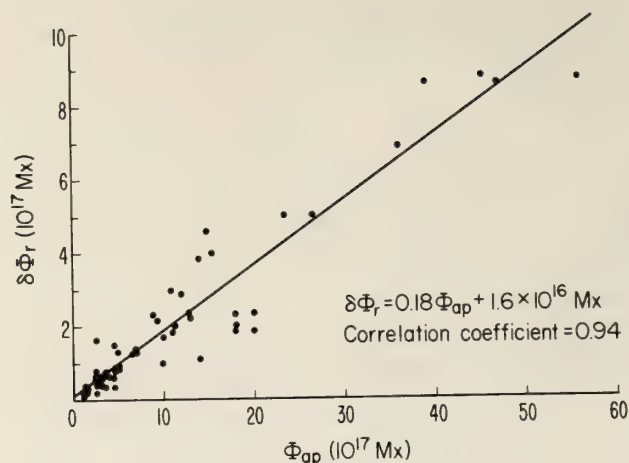


Figure 6. Comparison of the estimated returned magnetic flux as a function of the applied flux (Holzer and Slavin, 1979).

Since flux eroded from the dayside magnetosphere must pass across the polar cap to get to the night time magnetosphere, we would expect that the polar cap currents, which respond to the stresses imposed on the ionosphere by the flowing plasma at high altitudes, would be quite responsive to the interplanetary magnetic and/or electric field. Not only is there flow across the polar cap, but there are also changes in size of the polar cap as the imbalances develop between dayside and nightside reconnection. No attempt has been made to predict the size of the polar cap and test this prediction against auroral photos. However, an equivalent study of predicting the flux in the tail and comparing with the observed field strength has been performed successfully (Slavin and Holzer, 1979) as illustrated in Figure 7.

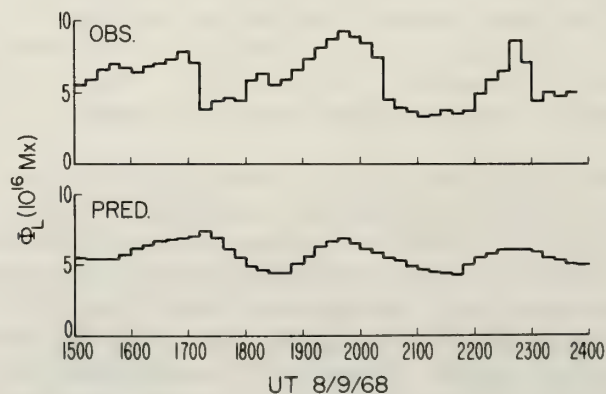


Figure 7. Predicted flux in the geomagnetic tail compared with the observed flux density using measured upstream solar wind conditions (Slavin and Holzer, 1979).

The polar cap equivalent current system is basically a double cell system called  $SP_q$  (Nagata and Kokubun, 1962). By comparing this current system with the IMF one finds that the polar cap current has an IMF dependent and an IMF independent part. The IMF dependent part is expected to have east-west asymmetries controlled by the  $B_y$  component of the interplanetary magnetic field. Convection should also be away from the sun unless the IMF is northward under which conditions merging could occur tailward of the cusp and cause sunward convection (Russell, 1972). Such effects have been observed (Friis-Christensen and Wilhjelm, 1975; Maezawa, 1976; Mishin, 1979). An attempt to predict such behavior in terms of a simple merging model in which merging takes place on the dayside only at places in which the magnetosheath and magnetospheric are antiparallel has been made by Crooker (1979). This model has been compared to the observations by Friis-Christensen (1979) and found to be in very good agreement.

#### 4. PREDICTIVE TECHNIQUES FOR GEOMAGNETIC INDICES

Geomagnetic indices are numbers which are designed to quantify the state of geomagnetic activity at a particular time. A description of the most commonly used indices is given as an appendix. There are high latitude indices such as  $K_p$ ,  $A_p$ ,  $a_a$ ,  $a_m$ , etc. which respond to auroral activity and distant magnetospheric currents and there is the low latitude Dst index which responds principally to the strength of the ring current or equivalently the energy stored in the radiation belts. If we can predict the magnitude of these indices then we feel we have at least an empirical understanding of the energy input to the magnetosphere. Furthermore, forecasts of the indices would permit users to plan programs or actions which take into account the expected level of geomagnetic activity.

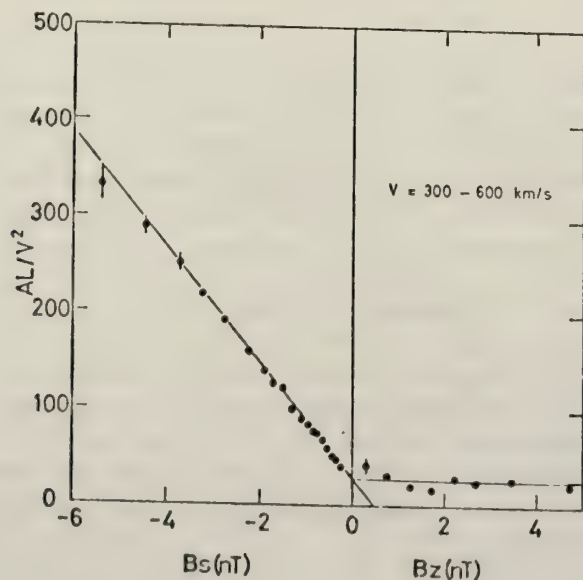


Figure 8. Normalized AL index as a function of the southward magnetic field. Negative values denote periods when interplanetary field was southward; positive values when it was northward (Murayama, 1979).

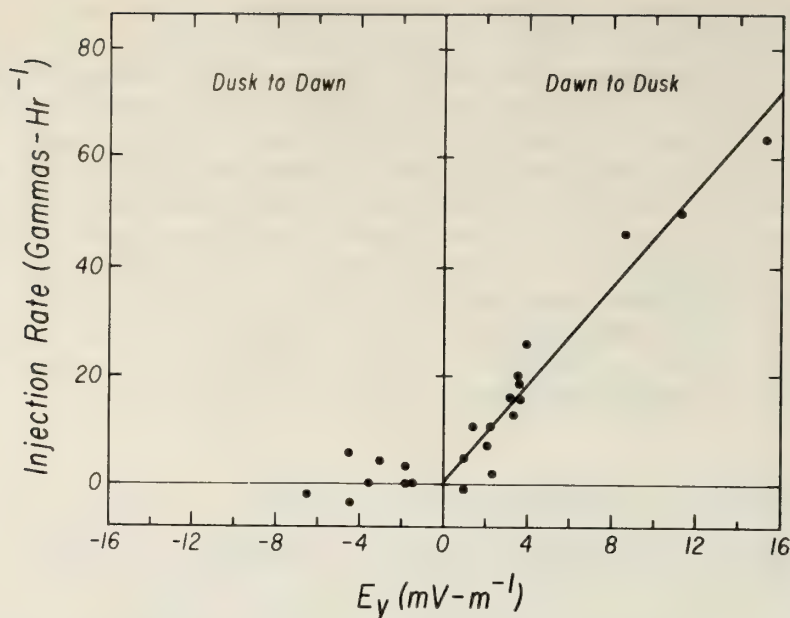


Figure 9. Injection rate into the ring current as a function of the solar wind dawn-dusk electric field,  $VB_z$  (Burton et al., 1975).

The AE index is the sum of the AU index and the AL index. Since statistical studies show these two indices appear to respond differently to the solar wind, we will treat them separately. As shown in Figure 8 for AL these indices appear to be linearly proportional to the southward field with little or no response for northward fields (Murayama, 1979; Maezawa, 1979) and not to  $B^2V \sin^4(\theta/2)$  hypothesized by Perreault (1976) and Akasofu (1978). This half-wave rectifier model has also been shown to be useful to predict Dst (Burton et al., 1975) as illustrated by Figure 9 and flux transport within the magnetosphere (Holzer and Slavin, 1978). The AL also is proportional to  $V^2$  as shown in Figure 10 (Murayama, 1979; Maezawa, 1979a). A similar result was found for Ap (Garrett et al., 1974; Crooker et al., 1977). The AU index is related to  $V$  rather than  $V^2$  (Maezawa, 1979a). Both AU and AL seem to depend to some extent on the variance of the interplanetary field and the solar wind density (Garrett, 1974; Maezawa, 1979b). This effect is particularly important during periods of northward interplanetary magnetic field (Maezawa, 1979b). One surprising result of the study of AL is that it "saturates" at about 500 gammas and is no longer linearly proportional to  $B_s V^2$  (Maezawa, 1979a). This could be due to the station distribution not being optimum during very disturbed conditions.

The Dst index can be predicted from the solar wind electric field allowing for the decay of energy from the ring current (Burton et al., 1975). In this model the half-wave rectifier model for the interaction was shown to be most appropriate and a linear dependence on  $V$  was assumed. Whether  $V^2$  would improve the model was not tested.



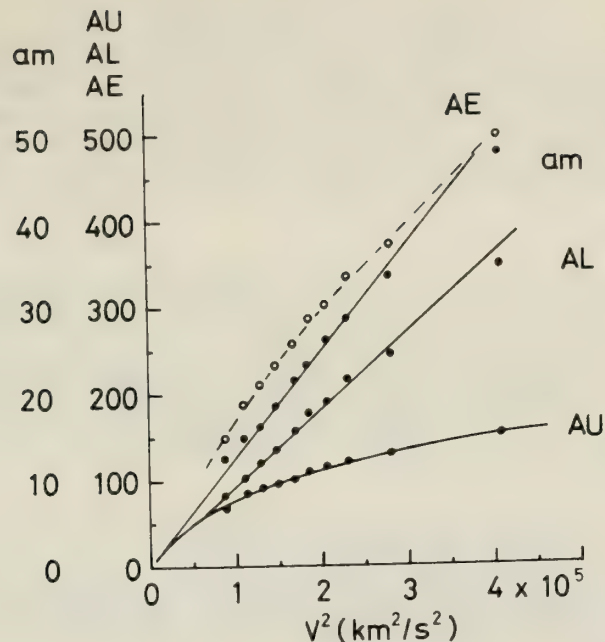


Figure 10. Dependence of various geomagnetic indices on the solar wind velocity (Maezawa, 1979).

As mentioned above  $A_p$  has been shown to be dependent on  $B_s V^2$ . This has been confirmed for  $a_m$  by Maezawa (1979a) (see figure 10) and for  $a_a$  by Feynman and Crooker (1978). The dependence of the midlatitude indices on density (Maezawa, 1979a) and on the variance of the interplanetary magnetic field (Garrett et al., 1974) is weak; however, it is more marked than that of the high latitude indices.

## 5. RECOMMENDATIONS

Although historically geomagnetic indices have been incorporated in many models predicting magnetospheric and ionospheric conditions, we now have the ability to characterize the state of the magnetosphere by more basic parameters. It is with this goal in mind that we recommend below the monitoring of a basic set of magnetospheric and interplanetary parameters that should significantly improve our forecasts. In some areas further study is recommended. This further study includes the investigation of possible future diagnostic techniques and basic research so that we understand the underlying physical mechanisms behind our present imperfect empirical knowledge. Finally, we recommend the continual computation, new computation, and in some cases tests of several solar and geophysical indices which will improve our predictive capabilities.

## 5.1 Spacecraft Measurements

a. Continuous in-situ measurements of the solar wind velocity number density and interplanetary magnetic field direction are required for near term geomagnetic forecasts. These measurements should be available in real time and should be taken as close to the earth-sun line as possible and preferably at a fixed distance from the earth such as at the forward libration point.

b. Since the magnetosphere responds to short time scale features ( $\sim 1/2$  hour) in the interplanetary field and since these features may have short coherence lengths in the solar wind we further recommend magnetic field measurements on a series of spacecraft in 20-40 Re geocentric orbit. These measurements will improve the accuracy of forecast of small to moderate disturbances, and should be available to forecasters in real time.

c. Since geomagnetic effects are first felt in the polar cap it is recommended that low altitude polar satellites be used to monitor the size of the polar cap, the width of the auroral oval, and the potential drop across the polar cap using measurements of particle fluxes, electric currents, and electric fields.

d. Historically auroral monitoring from the ground has provided retrospective information on the stage of substorm development. Spacecraft now can image an entire auroral oval from high apogee earth orbit. We recommend that such imaging systems be pursued and provide real time data to enable forecasts to follow the development of substorms.

e. Three dimensional pitch-angle measurements of  $E > 30$  kev electrons at synchronous orbit have shown their usefulness in predicting substorm occurrence. It is recommended that these data be made available in real time forecasters.

f. Present monitoring of the magnetic field at synchronous orbit by the SMS and GOES spacecraft has been impaired by the close proximity of the sensors to the spacecraft. We recommend that these sensors be so situated on future spacecraft that the full potential of these sensors for prediction purposes be realized.

g. Interplanetary scintillation measurements have not reached their full predictive potential because of the limited number of stations in operation around the world. New stations suitably placed would improve this situation.

h. As an aid in assessing the utility of coronal hole monitoring as a predictor of solar wind streams we urge the continuance of optical monitoring of the sun at  $10830 \text{ \AA}$  at least until full disk X-ray images from space become available.

## 5.2 Ground-Based Measurements

a. Currently, ground-based measurements (e.g. magnetic, etc.) also play a very important role in providing synoptic data which can be used for predictive purposes. We recommend that the quality and quantity of key ground-based measurements continue to be maintained, and improved if possible. (See recommendation (d) on Indices).

b. We recommend that new and better ways of collecting, using and distributing this data continue to be evolved.

c. The scientific results from new research facilities, such as the IMS North American Magnetometer Network, should be examined to see if they give clues to new indices or parameters which might eventually provide more effective predictions.

## 5.3 Future Studies

a. Present research has shown that geomagnetic activity depends on the solar wind velocity density and the interplanetary magnetic field. We recommend continued study of this dependence to quantify the functional dependence of the various magnetospheric processes on these parameters.

b. In view of the crucial role the geomagnetic tail plays in geomagnetic activity we urge the continued study of the tail and its responses to solar wind parameters.

c. Pc 3, 4 magnetic pulsations are highly correlated with interplanetary conditions but are not yet understood well enough to be used for diagnostic purposes. We recommend continued study of these pulsations to assess their diagnostic capability.

d. Kilometric radiation appears to be a sensitive indicator of the strength of auroral current systems. We urge that the diagnostic capability of kilometric radiation be assessed as its observation might provide a useful remote monitor of the auroral electrojet.

## 5.4 Indices

a. In view of the long history of the use of the Zurich sunspot number in geomagnetic studies we urge its continuance. The great strength of this data is that it forms a commensurate series.

b. The Comprehensive Flare Index has been shown to enhance significantly our ability to forecast geomagnetic activity. We recommend that it be routinely calculated in real time for all major flares.



c. The indices  $a_p$ ,  $a_m$  are apparently redundant. The relationships between them should be studied by the appropriate international body and one of them phased out. The better index should continue to be produced in a timely manner. A two station type aa index is simple to produce and may be useful in predictive applications.

d. The auroral electrojet indices AE, AU, and AL have shown great usefulness in retrospective studies of geomagnetic activity, but the delay in their preparation is unacceptably great, and the number and location of the stations limits their utility. We recommend that countries beneath the auroral oval insure continued monitoring of magnetic variations at auroral latitudes, install new sites where necessary for optimum station spacing, install digital recording and transfer data to the WDC's as rapidly as possible. Further we urge increased efforts to produce these indices on a more timely basis. Potential index applications for predictive uses might be met by prompt preliminary AE indices from a reduced network.

e. Data presently exist from several sources from which an index of polar size and auroral zone width might be obtained retrospectively. We urge that such indices be created and tested as indicators of the state of the magnetosphere and predictors of future conditions. These data should also be provided in real time to forecasters.

f. Data presently exist from which the hourly summations of southward magnetic field might be obtained. We recommend that such an index be prepared and issued in a form similar to the NSSDC Interplanetary Magnetic Field Data Book. We further recommend that data books containing the interplanetary magnetic field in solar magnetospheric coordinate and the solar wind number density and velocity be brought up to date and continue to be produced.

g. The ability to predict geomagnetic disturbances based on the 27-day recurrence of solar wind streams varies strongly across the solar cycle and from cycle to cycle. As an aid to planning a daily recurrence index should be developed and made available in real time.

h. Development of indices giving a global measure of the level of daytime pulsation activity in the Pc 3, 4 period range would serve both to advance the investigation of magnetospheric waves and to establish a quantitative basis for testing and improving the predictive capacity of pulsation observations. We recommend that such efforts be undertaken.

#### Acknowledgments

We would like to thank the following people who assisted us as corresponding members of our working group or who assisted us in our deliberations in Boulder: D. Baker, G.R. Heckman, P. Higbie, Y. Kamide, K. Marubashi, A. Nishida, G.L. Siscoe, V. Troitskaya and B. Tsurutani.

## REFERENCES

- Akasofu, S.-I. (1978): Interplanetary energy flux associated with magnetospheric substorms. Planet. Space Sci., 27, 425.
- Aubry, M.P., C.T. Russell, and M.G. Kivelson (1970): On the inward motion of the magnetopause preceding a substorm. J. Geophys. Res., 75, 7018.
- Baker, D.N., P.H. Higbie, E.W. Hones, Jr., and R.D. Belian (1978): High resolution energetic particle measurements at 6.6 Re. 3. Low-energy electron anisotropies and short-term substorm predictions. J. Geophys. Res., 83, 4863.
- Baker, D.N., P.R. Higbie, E.W. Hones, Jr., and R.D. Belian (1979): The use of >30 keV electron anisotropies at 6.6 Re to predict magnetospheric substorms, Proceedings of Solar Terrestrial Predictions Workshop.
- Blanc, M. (1978): Midlatitude convection electric fields and their relation to ring current development, Geophys. Res. Lett., 5, 203.
- Bolshakova, O.V., V.A. Troitskaya, and V.P. Hessler (1975): The diagnostics of the position of poleward boundary of the dayside cusp by means of intensity of high-latitude pulsations. Geomag. Aeron., 15, 755.
- Burch, J.L. (1972): Precipitation of low-energy electrons at high latitudes: Effects of interplanetary magnetic field and dipole tilt angle. J. Geophys. Res., 77, 6696.
- Burch, J.L. (1973): Rate of erosion of dayside magnetic flux based on a quantitative study of the dependence of polar cusp latitude on the interplanetary magnetic field. Radio Sci., 8, 955.
- Burton, R.K., R.L. McPherron and C.T. Russell (1975): An empirical relationship between interplanetary conditions and Dst. J. Geophys. Res., 80, 4204.
- Chang, S.C., and A. Nishida (1973): Spatial structure of transverse oscillations in the interplanetary magnetic field. Astrophys. Space Sci., 23, 301.
- Clauer, C.R., and R.L. McPherron (1979): Predicting partial ring current development, in Proceedings of International Solar Terrestrial Predictions Workshop.
- Crooker, N.U. (1979): Dayside merging and cusp geometry. J. Geophys. Res., submitted.
- Crooker, N.U., J. Feynmann, and J.T. Gosling (1977): On the high correlation between long-term averages of solar wind speed and geomagnetic activity. J. Geophys. Res., 82, 1933.



- Feynman, J., and N.U. Crooker (1978): The solar wind at the turn of the century. Nature, 275, 626.
- Friis-Christensen, E. (1979): The effect of the IMF on convection patterns and equivalent currents in the polar cap and cusp, in Magnetosphere Study 1979, 290, Japanese IMS Committee, Tokyo.
- Friis-Christensen, E., and J. Wilhjelm (1975): Polar cap currents for different directions of the interplanetary magnetic field in the Y-Z plane. J. Geophys. Res., 80, 1248.
- Garrett, H.B., T.W. Hill, and A.J. Dessler (1974): Influence of solar wind variability in geomagnetic activity. J. Geophys. Res., 79, 4603.
- Greenstadt, E.W. (1972): Field-determined oscillations in the magnetosheath as possible source of medium-period, daytime micropulsations, in Proc. of Conf. on Solar Terrestrial Relations, p. 515, Univ. of Calgary.
- Greenstadt, E.W., H.J. Singer, C.T. Russell, and J.V. Olson (1979): IMF orientation solar wind velocity and Pc 3-4 signals: A joint distribution. J. Geophys. Res., 84, 527.
- Gul'el'mi, A.V. (1974): Diagnostics of the magnetosphere and interplanetary medium by means of pulsations. Space Sci. Rev., 16, 333.
- Heppner, J.P. (1972): Polar cap electric field distributions related to the interplanetary magnetic field direction. J. Geophys. Res., 77, 4877.
- Holzer, R.E., and J.A. Slavin (1978): Magnetic flux transfer associated with expansions and contractions of the dayside magnetosphere. J. Geophys. Res., 83, 3831.
- Holzer, R.E., and J.A. Slavin (1979): A correlative study of magnetic flux transfer in the magnetosphere. J. Geophys. Res., in press.
- Horwitz, J.L., J.R. Doupnik, and P.M. Banks (1978): Chatanika radar observations of the latitudinal distribution of auroral zone electric fields, conductivities and fields. J. Geophys. Res., 83, 1463.
- Kamide, Y., and J.D. Winningham (1977): A statistical study of the 'instantaneous' nightside auroral oval: The equatorial boundary of electron precipitation as observed by the ISIS 1 and 2 satellites. J. Geophys. Res., 82, 5573.
- Maezawa, K. (1976): Magnetospheric convection induced by the positive and negative Z components of the interplanetary magnetic field: Quantitative analysis using polar cap magnetic records. J. Geophys. Res., 81, 2289.



- Maezawa, K. (1979a): Dependence of geomagnetic activity on the solar wind velocity and IMF parameters, in Magnetospheric Study 1979, 301, Japanese IMS Committee, Tokyo.
- Maezawa, K. (1979b): Statistical study of the dependence of geomagnetic activity on solar wind parameters, in Quantitative Models of the Magnetosphere, AGU Monograph Series, in press.
- Mishin, V.M. (1978): Electric fields and currents in the earth's magnetosphere, SibIZMIR, preprint.
- McPherron, R.L. (1970): Growth phase of magnetospheric substorms. J. Geophys. Res., 75, 5592.
- McPherron, R.L., C.T. Russell, and M.P. Aubry (1973): Satellite studies of magnetospheric substorms on August 15, 1968. 9. Phenomenological model of substorms. J. Geophys. Res., 78, 3131.
- Murayama, T. (1979): Principal factors controlling the development of the auroral electrojet, in Magnetospheric Study 1979, 296, Japanese IMS Committee, Tokyo.
- Nagata, T., and S. Kokubun (1962): Rep. Ionos. Space Res. Japan, 16, 2561.
- Orr, D. (1975): Probing the plasmopause by geomagnetic pulsations, Ann. Geophys., 31, 77.
- Orr, D. (1979): Geomagnetic pulsations and their relationship to structure within the magnetosphere, in Magnetospheric Study 1979, 124, Japanese IMS Committee.
- Perreault, P.D. (1976): On the relationship between IMF and magnetospheric storms and substorms. Ph.D. Thesis, University of Alaska.
- Perreault, P.D. and Y. Kamide (1976): A dawn-dusk asymmetry in the response of the magnetosphere to the IMF  $B_z$  component. J. Geophys. Res., 81, 4773.
- Pytte, T. (1978): Comments on 'A dusk-dawn asymmetry in the response of the magnetosphere to the IMF  $B_z$  component' by P.D. Perreault and Y. Kamide. J. Geophys. Res., 83, 2707.
- Russell, C.T. (1972): The configuration of the magnetosphere, in Critical Problems of Magnetospheric Physics, p. 1, IUCSTP Secretariat, National Academy of Sciences, Washington, D.C.
- Russell, C.T., and B.K. Fleming (1976): Magnetic pulsations as a probe of the interplanetary magnetic field: A test of the Borok B index, J. Geophys. Res., 81, 5882.

- Singer, H.J., C.T. Russell, M.G. Kivelson, E.W. Greenstadt, and J.V. Olson (1977): Evidence for the control of Pc 3, 4 magnetic pulsations by the solar wind velocity. Geophys. Res. Lett., 4, 377.
- Siscoe, G.L. (1979): A Dst contribution to the equatorward shift of the aurora. Planet. Space Sci., 26, in press.
- Slavin, J.A. and R.E. Holzer (1979): Empirical relationships between interplanetary conditions, magnetic flux transfer and the AL index, in Quantitative Modeling of Magnetospheric Processes, Geographical Monograph Series Vol. 21, (edited by W.P. Olson), in press, AGU, Washington, D. C.
- Southwood, D.J. (1968): The hydromagnetic instability of the magnetospheric boundary. Planet. Space Sci., 16, 587.
- Troitskaya, V.A., T.A. Pliasova-Bakunina, and A.V. Gul'el'mi (1971): Relationship between Pc 2-4 pulsations and the interplanetary magnetic field. Dokl. Akad. Nauk. SSSR, 197, 1312.
- Troitskaya, V.A., A.V. Gul'el'mi, O.V. Bolshakova, E.T. Matveyera, and R.V. Scheptnov (1972): Indices of geomagnetic pulsations. Planet. Space Sci., 20, 849.
- Wolfe, A., L.J. Lanzerotti, and C.G. MacLennan (1978): Dependence of hydromagnetic energy spectra on interplanetary parameters, (Abstract), EOS Trans. AGU, 59, 1166.
- Vero, J. (1975): Determination of the position of the magnetopause from geomagnetic pulsation indices. Acta. Geod. Geophys. Mont., 10, 459.

#### ADDRESSES OF SOLAR-WIND/MAGNETOSPHERE WORKING GROUP REPORT AUTHORS

C.T. Russell, Chairman  
Inst. of Geophys. and Planet. Phys.  
Univ. of Calif., Los Angeles, CA 90024

J.H. Allen  
WDC-A for Solar Terrestrial Phys.  
Environmental Data Services, NOAA  
Boulder, CO 80302

D.P. Cauffman  
National Aeronautics and Space Admin.  
Washington, D. C. 20546

J. Feynman  
Physics Department  
Boston College  
Chestnut Hill, MA 02167

E.W. Greenstadt  
TRW Systems Group  
Redondo Beach, CA 90278

R.E. Holzer  
S.M. Kaye  
J.A. Slavin  
Inst. of Geophys. and Planet. Phys.  
Univ. of Calif., Los Angeles, CA 90024

R.H. Manka  
National Science Foundation  
Washington, D. C. 20550

G. Rostoker  
Department of Physics  
University of Alberta  
Edmonton, Alberta, Canada

W.F. Stuart  
Inst. of Geological Sciences  
Murchison House  
Edinburgh EH 9 3LA  
Scotland

# ON THE PREDICTION OF MAGNETOSPHERIC CONFIGURATION

James A. Slavin

Robert E. Holzer

Department of Earth and Space Sciences  
Institute of Geophysics and Planetary Physics  
University of California  
Los Angeles, California 90024

A simple model of magnetic flux transfer is reviewed along with recently reported empirical expressions for flux transfer rates based upon the auroral AL index and the eastward component of the interplanetary electric field. It is shown that the magnitude of the time integral of the AL index over isolated intervals of geomagnetic disturbance can be predicted from the observed flux of southward interplanetary magnetic field lines incident on the front of the magnetosphere. Difficulties in the quantitative prediction of substorm onset time and intensity are discussed and a magnetospheric stress index based on both solar wind and ground based observations is proposed for this purpose. Finally, caveats concerning the use of both the AL index and interplanetary medium data in predicting the response of the magnetosphere to solar wind conditions are considered.

## Introduction

On time scales of minutes to days the configuration of the magnetosphere has been found to be variable with changes in solar wind stand-off distance, polar cusp latitude, polar cap area, magnetotail diameter and fields, and other features being correlated with both interplanetary conditions and geomagnetic activity (Burch, 1974). The ability to predict magnetospheric configuration from interplanetary and ground based observations is of particular interest in the formulation of dynamic models of the magnetosphere, an important step toward making quantitative solar-terrestrial forecasts. When the interplanetary magnetic field, or IMF, is southward, the rate of energy transfer from the solar wind to the magnetosphere is enhanced due to merging between the geomagnetic and solar wind magnetic fields (Dungey, 1961). Part of this energy is eventually dissipated in the ionosphere by particle precipitation and joule heating due to magnetospherically driven ionospheric currents associated with steady convection, substorms, and storms (Atkinson, 1967; Perreault, 1974) while the remaining energy flows down the tail to be returned to the solar wind. Accordingly, it may be possible to predict energy input to the magnetosphere from solar wind data and energy dissipation from ground magnetometer observations (e.g. Siscoe and Crooker, 1974; Burton et al, 1975). As these two energy rates will not always balance at any given time due to the variability of interplanetary conditions, the finite response time of the



magnetosphere to these changes, and the sporadic nature of substorms, the energy "state" of the magnetosphere changes with time. In the absence of large storm time ring currents this energy is to first order magnetic (Caan et al, 1973) with higher energy states corresponding to greater distortions of the earth's dipole field by increasing the amount of open flux in the tail (Russell and McPherron, 1973). For this reason it has been found most convenient to study changes in the state or configuration of the magnetosphere in terms of magnetic flux transfer (Levy et al, 1964; Russell, 1972; Burch, 1973).

In this paper the observed effects of magnetic flux transfer are briefly reviewed. Empirical relationships for the rates of flux transfer are examined for the purposes of predicting the time integrated magnitude of the AL index from solar wind observations. In addition, a magnetospheric stress index based on these empirical expressions for flux transfer is proposed for making short term predictions of substorm activity.

### Magnetic Flux Transfer

Net transfers of magnetic flux between the "closed" and "open" portions of the magnetosphere have been inferred from changes in the position of the forward magnetopause (Aubry et al, 1970; Holzer and Slavin, 1978a), variations in the size of the polar cap and cusp latitude (Burch, 1973; Akasofu and Kamide, 1976), observations of the tail magnetic field structure (Caan et al, 1973; Maezawa, 1975), and most recently by the in situ identification with ISEE 1 & 2 magnetometer data of reconnected flux tubes being swept tailward (Russell and Elphic, 1978). Based on these observations it has also been found that the magnetosphere possesses a minimum energy, or ground, state into which it relaxes during extended intervals of northward IMF (Akasofu and Kamide, 1976; Holzer and Slavin, 1978a).

The rate of erosion (i.e. the transfer of flux to the magnetotail) is controlled largely by solar wind conditions while the rate of flux return from the tail is determined by magnetospheric processes. Net transfers of flux,  $\Delta\Phi$ , result from imbalances between the two rates of transfer

$$\Delta\Phi = |\delta\Phi_e - \delta\Phi_r| \quad (1)$$

where  $\delta\Phi_e$  and  $\delta\Phi_r$  represent the total flux eroded and returned, respectively, in a given interval of time (Coroniti and Kennel, 1973; Holzer and Reid, 1975). When the flux being eroded exceeds that being returned, a contraction of the dayside magnetosphere takes place with the equilibrium position located nearer to the earth than in the case of a ground state configuration. During intervals with return exceeding erosion, the magnetopause is displaced outward toward its uneroded position. Under certain favorable circumstances these expansions and contractions are observed by satellites as multiple magnetopause crossings from which the net flux transfer can be calculated (Aubry et al, 1970). Solar wind dynamic pressure and/or magnetic field intensity just inside the magnetopause is used to separate the effects of compressions or rarefactions due to changes in external pressure from contractions and expansions resulting from alterations in configuration. Nine expansion and ten contraction events were found in 1968-9 Ogo 5 data by Holzer and Slavin (1978a) for which  $\Delta\Phi$  was inferred and the separate contributions of erosion and return identified (Holzer and Slavin, 1978b). The flux returned was found to be correlated with the time integral of the

AL index (Allen et al, 1968,1969) over the temporal extent of the event

$$\delta\Phi_r = 1.1 \times 10^{12} \int AL dt (\gamma - \min) + 4.0 \times 10^{15} \text{ Mx} \quad (2)$$

or in differential form

$$d\Phi_r/dt = 1.1 \times 10^{12} AL \text{ Mx/min} \quad (3)$$

where AL is taken to be a positive quantity measured in gammas (Holzer and Slavin, 1978b). This empirical relationship suggests that the maximum current density in the westward electrojet is proportional to the rate of return of flux from the tail. Such a finding is consistent with the various current disruption models of tail relaxation initiated by Atkinson (1967). It was also found by Holzer and Slavin (1978b) that the flux eroded was proportional to the flux of southward IMF impinging on the forward magnetosphere as would be expected from magnetic merging models (Levy et al, 1964)

$$d\Phi_e/dt = 0.2 d\Phi_{ap}/dt \quad (4)$$

where the applied flux of southward IMF,  $\Phi_{ap}$ , is

$$\Phi_{ap} = \int B_z^- V_{sw} W dt \quad (5)$$

In this expression the width of the magnetosphere is taken to be  $30R_e$ ,  $V_{sw}$  is the solar wind velocity, and  $B_z^-$  is the southward component of the IMF as determined from six minute, or shorter, averages of  $B_z$  in GSM coordinates. The more common use of hourly averages of  $B_z$  in studies such as these can seriously underestimate  $\Phi_{ap}$  in so far as the IMF north-south polarity often changes with a frequency on the order of minutes (Holzer and Slavin, 1978b,c).

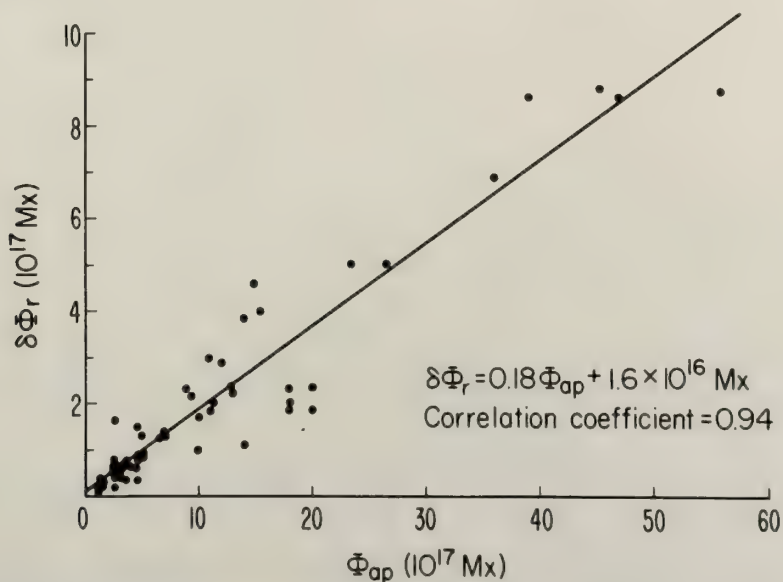


Fig. 1. The flux returned in each isolated event as inferred from the AL index and eq. 2 is plotted against the applied flux of southward interplanetary magnetic field.



The finding that the magnetosphere has a ground state is of some interest in that temporally isolated periods of geomagnetic activity will begin and end with the magnetosphere in this minimum energy configuration. Accordingly, this requires the total flux eroded and returned over the isolated interval to be equal. The flux returned as inferred from the AL index and equation (2) should then be proportional to  $\Phi_{ap}$  with a coefficient of  $0.2 \pm 0.1$ . Holzer and Slavin (1978b,c) have identified approximately 100 such isolated intervals in the first several months of 1967 and 1969 when satellite observations of the interplanetary medium were readily available. Figure 1 plots  $\delta\Phi_r$  versus  $\Phi_{ap}$  for the 56 periods between 1/1 and 3/31 of 1969. When more than one satellite was recording in the solar wind the average value of the applied flux is used. As shown the erosion efficiency is 18% and the correlation coefficient is 94%. The other periods considered with other spacecraft produced similar results. Hence, it is concluded that at least on a statistical basis geomagnetic activity as observed with the AL index can be predicted from interplanetary observations. Burton et al (1975) found this also to be the case for activity observed with the mid-latitude Dst index.

### Magnetospheric Stress Index

While the existence of the magnetospheric ground state enables the AL index to be quantitatively predicted on time scales of hours, forecasts on shorter time scales are not yet accurate due to a lack of understanding of magnetospheric processes such as substorm triggering (e.g. Kokubun et al, 1977). On both observational and theoretical grounds it has been found that substorm onset and intensity are related to both external solar wind conditions and previous magnetospheric activity (McPherron, 1972; Burch, 1972, Coroniti and Kennel, 1972). Slavin and Holzer (1978) have used the previously discussed relations for flux transfer rates to predict tail lobe flux during an interval of substorm activity when Ogo 5 was inbound in the tail near the noon-midnight meridian at  $X_{GSM} = -16$  to  $-10R_e$ . In Figure 2 the upper plot displays the observed lobe flux,  $\Phi_L$ , assumed to be  $\frac{TB}{2R} X_{GSM}^2 R_T^2$  with the tail radius constant at  $20R_e$ . While the position of the tail boundary and hence the magnetic fields are known to be variable (Maezawa, 1975), the error involved in assuming static cylindrical geometry for the near earth tail is not expected to be large. Substorm onset times determined from ground based magnetometers by Pytte and West (1978) are shown as solid vertical lines. It must also be noted that the sudden drop in  $\Phi_L$  following the onset of each substorm is diamagnetic in nature due to the expansion of the plasma sheet over the satellite (Russell et al, 1971). Accordingly, the intervals when particle data indicate that Ogo 5 was in the plasma sheet will not be considered,  $\sim 1705$  to  $\sim 1800$  and  $\sim 2015$  to  $\sim 2230$ . Displayed in the lower plot is the predicted lobe flux from net flux transfers calculated from the AL index, Explorer 34 observations of the interplanetary medium, and equations (3) and (4) in 12 minute averages with the initial value of  $\Phi_L$  set equal to that observed by Ogo 5 at 1500. In addition, as equations (3) and (4) apply to the forward magnetosphere a 24 minute shift of the predicted flux toward earlier times has been made to allow for transfer to the tail (e.g. Rostoker et al, 1972). The agreement between observation and prediction is good considering the simple model assumed. Further study of the usefulness of this method of predicting tail lobe flux in the absence of in situ data is continuing at this time.



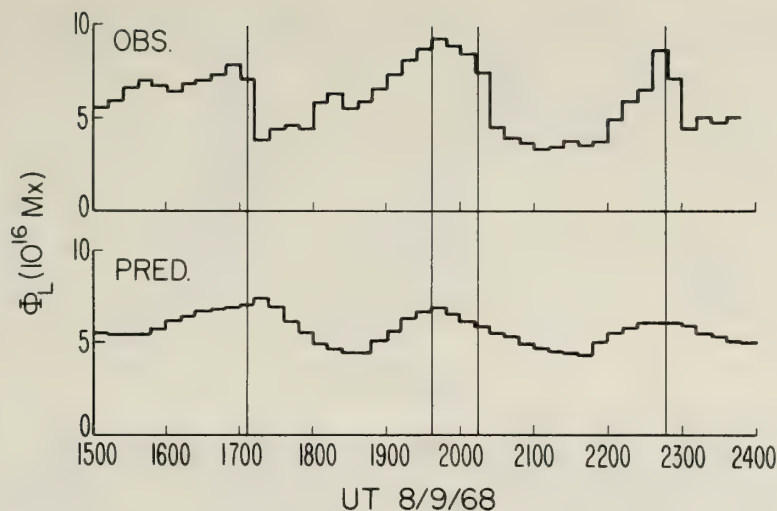


Fig. 2. The upper plot displays the tail flux calculated from Ogo 5 magnetic observations in the magnetotail while lower plot depicts the flux predicted from the interplanetary and ground data.

Due to the association between substorm occurrence and intensity and the net transfer of flux to the magnetotail (Caan et al, 1977, 1978) it may be desirable for predictive purposes to create a magnetospheric stress index indicative of the net flux transfer occurring. Based on equations (3) and (4) we propose an hourly index, T to T+60 min,

$$\begin{aligned}
 \text{MS}(T, T+60\text{min}) &= \int \left( \frac{d\Phi}{dt} e - \frac{d\Phi}{dt} r \right) dt \\
 &= \int_{T-20\text{min}}^{T+40\text{min}} (aV_{\text{sw}} B_z^- - bAL) dt
 \end{aligned} \tag{6}$$

where solar wind velocity is measured in km/sec,  $B_z^-$  in  $\gamma$ 's, AL in positive  $\gamma$ 's,  $a=2.3 \times 10^{11}$ ,  $b=1.1 \times 10^{12}$ , and time in minutes. Positive values of MS such as 1500 to 1700, 1800 to 2000, and 2100 to 2300 in Figure 2 correspond to enhancements of tail stress while negative MS values as is the case for 1700 to 1800, 2000 to 2100, and 2300 to 2400 correspond to a relaxation of the tail predominantly through the substorm process in this example. Recently, particle pitch angle distributions have been used for similar purposes by Pytte and West (1978), West et al (1978), and Baker et al (1978). Both approaches are likely to produce useful substorm diagnostics both for predictive and research purposes.

#### Measurement of $\Phi_{\text{ap}}$ and AL

All methods of predicting magnetospheric parameters from interplanetary data must ultimately depend to some extent on the ability to quantitatively assess conditions both within and without the magnetosphere with some degree

of precision. In this paper it has been  $V_{sw}$ ,  $B_z^-$ , and the westward electrojet that are considered. Solar wind velocity measured simultaneously by more than one near earth satellite does not usually differ greatly and solar wind dynamic pressure from these satellites can be used as an excellent predictor of uneroded magnetopause position (Fairfield, 1971; Holzer and Slavin, 1978a). However, the observed values of  $B_z^-$  can vary greatly among even satellites separated by only tens of earth radii and hence may not be uniform over the front of the magnetosphere (e.g. Perreault and Kamide, 1978; Pytte, 1978). Figure 3 plots  $\Phi_{ap}$  inferred from Explorer 35 observations against those from Heos I for the 19 out of the 56 points in Figure 1 when both satellites were making observations in the solar wind. In the individual cases the values of  $B_z^-$  measured differ greatly and in a manner that is random and cannot be attributed magnetometer offsets. The correlation coefficient is a low 71% for dependence of the applied flux measured at one satellite on that observed at the other. For this reason ISEE 3 data gathered from a halo orbit about a lagrangian point  $250R_e$  upstream should be compared with observations made by earth orbiting satellites and geomagnetic activity to assess its usefulness in inferring the magnetospheric boundary conditions. Such a study is particularly timely in view of the anticipated role of an interplanetary laboratory satellite in an ISEE 3 type orbit as part of the proposed "OPEN" mission (Cauffman, 1978).

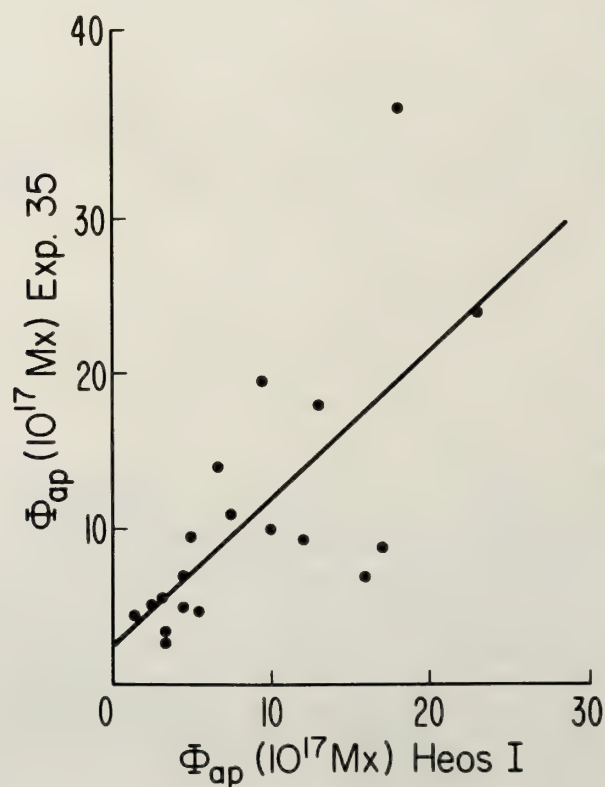


Fig. 3. For each of the 19 events from the first figure that had data available from both Explorer 35 and Heos I the applied flux seen at one is plotted against that observed at the other.

The AL index was first introduced by Davis and Sugiura (1966) as a measure of the maximum westward electrojet current density as a function of time. Ideally, a large number of observing stations over a wide range in both latitude and longitude would be used for this purpose. In practice, the auroral indices are calculated from data provided by 11 stations located from about  $60^{\circ}$  to  $70^{\circ}$  in geomagnetic latitude and nonuniformly distributed in longitude. As the position of the electrojets is variable to the extent that they are not always near one of these observatories the AL index does not always fulfill its intended purpose (e.g. Holzer and Slavin, 1978b). Further discussion of the characteristics and uses of this index can be found in Rostoker (1972), Garrett (1973), Perreault (1974), and Kamide (1978).

### Conclusions

At this time the prediction of both magnetospheric configuration and activity is limited by the means available to observe interplanetary conditions and the magnetospheric response. However, it is clear that on a statistical basis both long and short term forecasts can now be made with some degree of confidence. Future advances will depend in part on the creation of quantitative nonstatic models of the magnetosphere which are functions of the solar wind conditions.

### Acknowledgements

The authors are pleased to acknowledge the use of interplanetary medium data sets prepared by the space physics group at the UCLA Institute of Geophysics and Planetary Physics. We are also indebted to the co-investigators, Paul J. Coleman, Jr., Thomas A. Farley, and Darrell L. Judge for providing us with Ogo 5 fluxgate magnetometer data. The research reported in this paper was supported by National Science Foundation grant ATM 75-01431 and National Aeronautics and Space Administration grant NGR 05-007-276. Institute of Geophysics and Planetary Physics publication 1891.

### References

- Akasofu, S.-I., and Y. Kamide (1976): Substorm energy, Planet. Space Sci., 24, 223.
- Allen, J.H., C.C. Abston, and L.D. Morris (1968): Auroral Electrojet Magnetic Activity Indices, World Data Center Report UAG-29, Boulder, Colorado.
- Allen, J.H., C.C. Abston, and L.D. Morris (1969): Auroral Electrojet Magnetic Activity Indices, World Data Center Report UAG-31, Boulder, Colorado.
- Atkinson, G. (1967): An approximate flow equation for geomagnetic flux tubes and its application to the polar substorm, J. Geophys. Res., 72, 5373.
- Aubry, M.P., C.T. Russell, and M.G. Kivelson (1970): Inward motion of the magnetopause before a substorm, J. Geophys. Res., 75, 7018.
- Baker, D.N., P.R. Higbie, E.W. Hones, Jr., and R.D. Belian (1978): High resolution energetic particle measurements at  $6.6R_E$ : 3. Low energy electron anisotropies and short-term substorm predictions, J. Geophys. Res., 83, 4863.



- Burch, J.L. (1972): Preconditions for the triggering of polar magnetic substorms by storm sudden commencements, J. Geophys. Res., 77, 5629.
- Burch, J.L. (1973): Rate of erosion of dayside magnetic flux based on a quantitative study of the dependence of the polar cusp latitude on the IMF, Radio Sci., 8, 955, 1973.
- Burch, J.L. (1974): Observations of interactions between interplanetary and geomagnetic fields, Rev. Geophys. Space Phys., 12, 363.
- Burton, R.K., R.L. McPherron, and C.T. Russell (1975): An empirical relationship between interplanetary conditions and Dst, J. Geophys. Res., 80, 4204.
- Caan, M.N., R.L. McPherron, and C.T. Russell (1973): Solar wind and substorm-related changes in the lobes of the geomagnetic tail, J. Geophys. Res., 78, 8087.
- Caan, M.N., R.L. McPherron, and C.T. Russell (1977): Characteristics of the association between the IMF and substorms, J. Geophys. Res., 82, 4837.
- Caan, M.N., R.L. McPherron, and C.T. Russell (1978): The statistical magnetic signature of magnetospheric substorms, Planet. Space Sci., 26, 269.
- Cauffman, D.P. (1978): The OPEN program: A major opportunity for future space plasma research, Final Program of the AGU Chapman Conference on Magnetospheric Substorms and Related Plasma Processes, Los Alamos Scientific Laboratory, Los Alamos, New Mexico, October 9-13.
- Coroniti, F.V., and C.F. Kennel (1972): Changes in magnetospheric configuration during the substorm growth phase, J. Geophys. Res., 77, 3361.
- Coroniti, F.V., and C.F. Kennel (1973): Can the ionosphere regulate magnetospheric convection?, J. Geophys. Res., 78, 2837.
- Davis, T.N., and M. Sugiura (1966): Auroral electrojet activity index and its universal time variations, J. Geophys. Res., 71, 785.
- Dungey, J.W. (1961): Interplanetary magnetic field and the auroral zones, Phys. Rev. Letts., 6, 37.
- Fairfield, D.H. (1971): Average and unusual locations of the earth's magnetopause and bow shock, J. Geophys. Res., 76, 6700.
- Garrett, H.B. (1973): The role of fluctuations in the IMF in determining the magnitude of substorm activity, Planet. Space Sci., 22, 111.
- Holzer, R.E., and G.C. Reid (1975): The response of the dayside magnetosphere-ionosphere system to time-varying field line reconnection at the magnetopause, J. Geophys. Res., 80, 2031.
- Holzer, R.E., and J.A. Slavin (1978a): Magnetic flux transfer associated with expansions and contractions of the dayside magnetosphere, J. Geophys. Res., 83, 3831.

- Holzer, R.E. and J.A. Slavin (1978b): A correlative study of magnetic flux transfer in the magnetosphere, In press, J. Geophys. Res.
- Holzer, R.E., and J.A. Slavin (1978c): Magnetospheric energy flux during solar cycle 20 (abstract), EOS Trans. Am. Geophys. Union, 59, 1167.
- Kamide, Y. (1978): On the auroral electrojet indices, preprint, Kyoto Industrial University, Kyoto, Japan, 1978.
- Kokubun, S., R.L. McPherron, and C.T. Russell (1977): Triggering of substorms by solar wind discontinuities, J. Geophys. Res., 82, 74.
- Levy, R.H., H.E. Petschek, and G.L. Siscoe (1964): Aerodynamic aspects of the magnetospheric flow, AIAA J., 2, 2065.
- Maezawa, K. (1975): Magnetotail boundary motion associated with geomagnetic substorms, J. Geophys. Res., 80, 3543.
- McPherron, R.L. (1972): Substorm related changes in the geomagnetic tail: The growth phase, Planet. Space Sci., 20, 1521.
- Perreault, P.D. (1974): On the relationship between IMF and magnetospheric storms and substorms, Ph.D. thesis, University of Alaska, 1974.
- Perreault, P.D., and Y. Kamide (1978): Reply, J. Geophys. Res., 83, 2709.
- Pytte, T. (1978): Comment on "A dusk-dawn asymmetry in the response of the magnetosphere to the IMF  $B_z$  components", by P.D. Perreault and Y. Kamide, J. Geophys. Res., 83, 2707.
- Pytte, T., and H.I. West (1978): Ground-satellite correlations during pre-substorm magnetic field configuration changes and plasma sheet thinning in the near-earth magnetotail, J. Geophys. Res., 83, 3791.
- Rostoker, G. (1972): Geomagnetic indices, Rev. Geophys. Space Phys., 10, 935.
- Rostoker, G., H.-L. Lam, and W.D. Humer (1972): Response time of the magnetosphere to the interplanetary electric field, Can. J. Phys., 50, 544.
- Russell, C.T. (1972): The configuration of the magnetosphere, Critical Problems of the Magnetospheric Physics (Proc. of COSPAR/IAGA/URSI symposium, Madrid, Spain) edited by E.R. Dryer.
- Russell, C.T., and R.C. Elphic (1978): ISEE observations of flux transfer events at the dayside magnetopause, submitted to Geophys. Res. Lett.
- Russell, C.T., and R.L. McPherron (1973): The magnetotail and substorms, Space Sci. Rev., 15, 205.
- Russell, C.T., R.L. McPherron, and P.J. Coleman, Jr. (1971): Magnetic field variations in the near geomagnetic tail associated with weak substorm activity, J. Geophys. Res., 76, 1823.

- Siscoe, G.L., and N.U. Crooker (1974): A theoretical relation between Dst and the solar wind merging electric field, Geophys. Res. Letts., 1, 17.
- Slavin, J.A., and R.E. Holzer (1979): Empirical relationships between interplanetary conditions, magnetospheric flux transfer, and the AL index, to appear in Quantitative Modeling of Magnetospheric Processes, Geophysical Monograph Series, vol. 21, edited by W.P. Olson, AGU, Washington, D.C.
- West, H.I., Jr., R.M. Buck, and M.G. Kivelson (1978): On the configuration of the magnetotail near midnight during quiet and weakly disturbed period: State of the magnetosphere, J. Geophys. Res., 83, 3805.



# THE MAGNETOSPHERIC ELECTRIC FIELD AND CONVECTIVE PROCESSES AS DIAGNOSTICS OF THE IMF AND SOLAR WIND

Stanley M. Kaye  
Department of Earth and Space Sciences  
and  
Institute of Geophysics and Planetary Physics, University of California  
Los Angeles, California 90024 U.S.A.

The convection electric field in the equatorial plane is a poorly determined magnetospheric feature due to the high degree of difficulty in measuring its rather low, typically  $< 1 \frac{\text{mV}}{\text{m}}$ , magnitude. Consequently, it is a poor diagnostic of solar wind and IMF conditions. However, indirect measurements of the convection field as well as direct or indirect measurements of the ionospheric electric field provide a means to at least monitor qualitatively solar wind processes. For instance, asymmetries in the ionospheric electric field and ionospheric Hall currents over the polar cap reflect the solar wind sector polarity. A stronger electric field, and thus convective flow, is found on the side of the polar cap where the y component of the IMF is parallel to the y component of the geomagnetic field. Additionally, the magnitude of the electric field and convective processes at low altitudes and in space increases with increasing southward  $B_z$  and/or solar wind velocity, and thus may indicate the arrival at earth of an interaction region in the solar wind. It is apparent that processes associated with the convection electric field may be used to predict large-scale features in the solar wind; however, with our present empirical knowledge it is not possible to make quantitative predictions of individual solar wind or IMF parameters.

## 1. INTRODUCTION

Because of limitations in instrument sensitivity, efforts to measure directly the earth's electric field have been most successful at low altitudes where typical field magnitudes lie in the range 10 to 60  $\frac{\text{mV}}{\text{m}}$ , and least successful in space where the intensities are typically  $< 1 \frac{\text{mV}}{\text{m}}$ . Furthermore, the mapping of the electric field from low altitudes into the magnetosphere is complicated by our not yet being able to predict where, when, and how large a potential drop parallel to the magnetic field will exist. Consequently, we have a very poor empirical picture of the convection electric field in the equatorial magnetosphere.

It would therefore seem rather implausible to use the convection electric field as a diagnostic of the solar wind and interplanetary magnetic field (IMF) unless there are effects associated with the convection field which are measurable and which can yield the desired information. Fortunately such

features exist. Although not a one-to-one mapping of the convection field, the ionospheric electric field reflects its gross changes in magnitude and direction. These ionospheric fields may be measured directly by satellites, balloons, and rockets, and indirectly by radar measured drift velocities or ground magnetic perturbations caused by convection induced ionospheric and magnetospheric currents.

These techniques give us a way to infer the nature and magnitude of the convection field; we might now ask what aspects of the solar wind and IMF may be monitored with the magnetospheric information. This question necessarily implies a causal relation between the interplanetary medium and magnetosphere, and to answer it requires some understanding of the interplanetary-magnetospheric coupling mechanism. It is therefore worthwhile to briefly consider two possible coupling mechanisms.

In the open model of the magnetosphere (Dungey, 1961; Coleman, 1971) we find a direct coupling between geomagnetic and interplanetary field lines within some spatially limited region. In this model, the convection electric field is proportional to the dawn-dusk component of the interplanetary electric field (IEF), with the IEF in the collisionless solar wind plasma given by  $\vec{E} = -\vec{V}_{SW} \times \vec{B}$  where  $\vec{V}_{SW}$  is the solar wind velocity and  $\vec{B}$  is the total interplanetary magnetic field vector. A southward component of the IMF gives rise to a dawn-dusk directed component of the IEF. Furthermore, where the IMF couples to the geomagnetic field may be influenced by the  $B_y$  component of the IMF ( $B_y > 0$  indicating a dawn-dusk directed component), and thus asymmetries in magnetospheric convective flow may result if  $B_y \neq 0$ .

On the other hand, the IMF dependence is not present in the closed model of the magnetosphere (Axford and Hines, 1961) in which the geomagnetic field is completely detached from the IMF. Here, the magnetospheric convection field still depends on the momentum transfer from solar wind to magnetosphere, and in a frictional drag model a greater transfer of momentum comes about with increased solar wind velocity.

So the convection field may depend on both the IMF and solar wind velocity; the natural question to ask, then, is whether any particular IMF and/or velocity characteristics are associated with macroscopic solar wind structures. Clearly, the answer is yes, with the most prominent feature being the solar wind sector structure which controls the polarity of  $B_y$ . We find that dawn-dusk asymmetries in convective processes over the polar cap depend very strongly on the magnitude and direction of  $B_y$ , and therefore these asymmetries may be used to infer sector polarity.<sup>y</sup> In a recent study, Rosenberg and Coleman (1978) showed that for the current dipolar-solar cycle large southward  $B_z$  and enhanced solar wind velocity are characteristic of the leading edge of a solar wind interaction region, a region of interaction between solar wind low and high velocity streams. The enhancement of magnetospheric convection processes may therefore indicate the arrival at earth of such a region. As we shall see, however, we are not yet at the stage where we can use magnetospheric convection to infer microscopic characteristics (i.e., magnitudes of the individual parameters) of the solar wind.

Now that we have some foundation of knowing the types of solar wind structures and parameters that can be monitored by the convection field, and how the convection field itself can be monitored, we turn to the observations. I have structured this report in the format of a review paper since some indication of the empirical relation between convection processes and interplanetary conditions may be derived from previously published work.



Studies concerning the relation of the convection electric field and convection processes fall into two main categories; those performed at high latitudes (polar cap), and those done at low latitudes (auroral zone and below). These categories may be subdivided into types of studies (i.e., direct electric field measurement vs. inference of electric field from measurements of plasma drifts, ground magnetic perturbation, particle intensities, or auroral forms). I have adopted such a classification scheme for this report. Since I have concentrated on reporting the more recent work on the interplanetary-convection electric field relation, I refer the reader to the work of Nishida (1975), Fairfield (1977), and Stern (1978) for a more comprehensive review of electric field measurements.

## 2. HIGH LATITUDE MEASUREMENTS (POLAR CAP)

At high latitudes direct electric field measurements were obtained by both satellites and balloons. Heppner (1972a,b) studied the convection electric field from measurements taken aboard the low-altitude ( $\sim 400$  km) polar-orbiting OGO-6 satellite. In the former paper Heppner found the polar cap electric field to be directed dawn-dusk for all levels of geomagnetic activity, and in the latter paper he showed that the polar cap field exhibited a strong systematic variation in the dawn-dusk meridian which depended on  $B_y$ . Heppner found a stronger electric field, or equivalently faster convection, on the side of the polar cap where the  $y$  component of the IMF was parallel to the  $y$  component of the geomagnetic field. To be specific, he found that the electric field in the northern polar cap was stronger at dawn (dusk) for  $B_y > 0$  ( $B_y < 0$ ), and found the opposite behavior in the southern polar cap.

Measurements obtained from balloons launched at geomagnetic latitudes from  $\lambda = 78^\circ$  to  $\lambda = 87^\circ$  (Mozer et al., 1974) supported the  $B_y$  related dawn-dusk asymmetry of the polar cap convection electric field magnitude. Mozer et al. further determined an empirical relation between the magnitude of the average polar cap electric field and  $B_z$ , the north-south component of the IMF. They found that  $E(\frac{mV}{m}) = 22 - 3B_z(\gamma)^z$  where  $E > 0$  indicates a dawn-dusk directed convection field. A notable feature of this relation is the anti-sunward convection over the polar cap even in the presence of northward  $B_z$  ( $B_z > 0$ ). Mozer noted that the potential drop across the polar cap for  $B_z = 0$  is  $\sim 60$  kV, a value commensurate with the Heppner (1972a) measurements during times of little geomagnetic activity.

In more recent balloon-borne electric field measurements D'Angelo et al. (1976) studied the relation between the magnitude of the polar cap convection field and the earth's position within the solar sector structure. As the sector boundary passed the earth, the convection field increased by a factor of 2 to 3 from its pre-boundary passage value; the magnitude then decreased throughout the remainder of the sector. Bahnsen and D'Angelo (1976) found similar behavior in the IEF. Here, the IEF increased by a factor of 3 to 6 at the sector boundary (to a value of 3 to 6 mV/m), and they observed the increase at all sector boundaries regardless of whether the transition was from toward to away or vice-versa. Bahnsen and D'Angelo further showed the IEF variation to be due both to variations in the solar wind velocity and IMF.

Ground magnetic perturbations and the corresponding equivalent current systems have been extensively used to study convection and the driving electric field at high latitudes. Using this technique, Friis-Christensen and Wiljhelm (1975) determined that the circumpolar circulation in the northern



polar cap was eastward (westward) for  $B_y > 0$  ( $B_y < 0$ ). In a comprehensive study of polar cap magnetograms, Maezawa (1976)<sup>y</sup> found that strong convection in the northern hemisphere shifted towards dawn (dusk) for  $B_y > 2\gamma$  ( $B_y < -2\gamma$ ), but found no asymmetry for  $|B_y| < 2\gamma$ . Nishida (1975) pointed out that the circulation pattern found by Friis-Christensen and Wiljhelm (1975) would be observed when the average convection pattern ( $B_y \approx 0$ ) was shifted towards dawn (dusk) for  $B_y > 0$  ( $B_y < 0$ ).

Maezawa<sup>y</sup> (1976) additionally studied the relation between  $H_{\text{vert}}$  (the ground magnetic perturbation along the ambient field produced by ionospheric Hall currents) and  $B_z$ . Here, the magnitude of the Hall currents directly reflects the magnitude of the convective flow. For  $B_z < 0$ ,  $H_{\text{vert}}$  responded linearly to an increase in  $|B_z|$  in a manner consistent with increasing anti-sunward convection over the polar cap with increasing southward  $B_z$ . The day-side magnetospheric response occurred from 0 to 30 min. after the change in  $B_z$ . For  $B_z < -1\gamma$ , Maezawa also noted a linear relation between  $H_{\text{vert}}$  and solar wind velocity, with increasing anti-sunward convection corresponding to larger solar wind velocities. The unusual observation in Maezawa's study was the observed linear response between  $H_{\text{vert}}$  and  $B_z$  for  $B_z > 0$ , but here with the  $H_{\text{vert}}$  perturbation indicating sunward convection over the central polar cap. He suggested that reconnection between the northward IMF and lobe magnetic field on the tail magnetopause could account for such an observation. This suggestion was first put forward by Dungey (1961).

Using data from the low altitude polar-orbiting DMSP-32 satellite, Meng et al. (1977) correlated the direction of the dawn-dusk intensity gradient of low energy (200 eV) electrons in the polar cap with the direction of  $B_y$ . In 79 of 88 passes they found a factor of 2 to 5 particle flux increase which was directed dusk-dawn (dawn-dusk) over the northern (southern) polar cap for  $B_y > 0$ , and an oppositely directed decrease over the respective polar cap for  $B_y < 0$ . Meng et al. noted that the higher electron fluxes occurred on the side of the polar cap at which a stronger convection field was expected, and suggested that the flux gradient could result from non-uniform entry of electrons into the polar cap region.

D'Angelo and Oleson (1975) inferred the convection field magnitude in the polar cap from the occurrence of the "slant E condition" (SEC) over Godhavn, Greenland. SEC, caused by the Farley-Buneman instability in which the relative drift velocity between the electron and ion populations in the E-region exceeds the local sound speed, maximized within the first three days after a sector boundary passed the earth, and subsequently decreased with time towards the end of the sector. In this manner the frequency of SEC occurrence correlates well with the variation in the magnitude of the convection field reported by D'Angelo et al. (1976). Also similar to the electric field variation was equivalent frequency of occurrence of SEC regardless of whether the sector was towards or away. However, D'Angelo and Oleson (1975) noted that no SEC occurred for  $V_{\text{sw}} < 350$  to 400 km/sec. According to their calculations, the electric field near the magnetopause corresponding to  $V_{\text{sw}} = 400$  km/sec. maps to an E-region electric field of  $\sim 20$  mV. The authors note that this electric field value is near the threshold value for the Farley-Buneman instability to occur. That there is a direct empirical relation between the solar wind velocity and the magnitude of the convection electric field was first demonstrated by Vasyliunas (1968).

### 3. LOW LATITUDE MEASUREMENTS (BELOW POLAR CAP)

At lower latitudes the convection electric is found to respond to changes in the solar wind and IMF in much the same way as was observed in the polar cap. For instance, direct electric field measurements by balloons launched at northern geomagnetic latitudes corresponding to  $L = 5.4$  to  $8.2$  revealed stronger north-south electric fields at dusk for  $B_y < 0$  and at dawn for  $B_y > 0$  (Mozer and Lucht, 1974). From observations of auroral forms, Lassen and Danielson (1978) showed that the zone of quiet time particle precipitation in the northern hemisphere shifted towards dawn (dusk) for  $B_y > 0$  ( $B_y < 0$ ). The shift was most pronounced when  $B_z > 0$ .

Parks and Pellat (1972) found that 50-150 keV electron fluxes measured aboard the geosynchronous ATS-1 between 1400 and 2400 LT exhibited peaks (troughs) when  $B_z > 0$  ( $B_z < 0$ ), but found no flux modulation in either the 150-500 keV or 500-1000 keV channel. Since no solar wind pressure variations were observed at the times of the 50-150 keV flux modulations, they assumed that the peaks and troughs represented samplings from different portions of the spatial electron distribution. An electron population, with an earthward directed flux gradient, that penetrated farther earthward when  $B_z < 0$  (stronger convection electric field) and that shifted anti-earthward when  $B_z > 0$  (weaker convection electric field) would cause such a modulation. The authors suggested that the modulation was observed only in the 50-150 keV channel since the average drift period of these particles was similar to the  $\sim 100$  min. time scale of the  $B_z$  variation.

Measurements of  $>1.6$  MeV and  $>3.9$  MeV electrons aboard the geosynchronous ATS-6 satellite revealed a modulation of these particle fluxes by the solar wind sector structure (Paulikas and Blake, 1976). In the northern hemisphere the highest electron fluxes occurred when  $B_y < 0$  in the spring and when  $B_y > 0$  in the autumn. According to the authors, the higher level of fluxes indicated a greater transfer of energy from solar wind to the magnetosphere, and to support this claim they noted that the IMF would have a component anti-parallel to the geomagnetic field in the northern hemisphere for the above  $B_y$  and seasonal combinations.

Horwitz et al. (1978) used the Chatanika radar facility ( $\lambda = 63^\circ$  to  $68^\circ$ ) to monitor F-region convection as a function of  $B_z$  and found the inferred convection field magnitude to be proportional to this IMF component. They present one example in which the northward directed field at dusk increased from 0 to  $30-60 \frac{\text{mV}}{\text{m}}$  as  $B_z$  decreased from 0 to  $-5\gamma$ , and one in which the electric field increased to  $\sim 100 \frac{\text{mV}}{\text{m}}$  as  $B_z$  went from  $3\gamma$  to  $-10\gamma$ . In addition, they found a simultaneous response of the electric field at  $66.7^\circ$  to  $67.9^\circ$  to southward changes in  $B_z$ , but found a  $\sim 30$  min. delayed response at  $63.1^\circ$ , suggesting "movement of a belt of northward E equatorward into the Chatanika field of view in response to a southward IMF." Blanc (1978) studied the penetration of the convection field to much lower latitudes with the Saint-Santin radar facility ( $\lambda = 47^\circ$ ). Quiet time east-west drift velocities of  $\sim 10$  m/s ( $E \approx 0.4 \frac{\text{mV}}{\text{m}}$ ) increased to 50 to  $150 \frac{\text{m}}{\text{s}}$  ( $E \approx 2$  to  $6 \frac{\text{mV}}{\text{m}}$ ) within 30 min. after  $B_z$  turned and remained steadily southward at  $-7\gamma$ . He also showed that the rate of plasma injection, as measured by  $\Delta D_{st} / \Delta t$ , increased with increasing east-west drift velocity (increasing convection field strength). A prior theoretical study by Siscoe and Crooker (1974) and an empirical study by Burton et al. (1975) showed that the rate of plasma injection was proportional to  $E_y$ . The latter study determined the time delay to be  $25 \pm 40$  min.



Finally, using a chain of midlatitude magnetic observatories, Clauer and McPherron (1978) showed that the strength of the partial ring current, and thus the magnitude of the convection field causing such plasma penetration, was proportional to the magnitude of southward  $B_z$ . Within an hour after  $B_z$  turned and remained steadily southward at  $\sim 5\gamma$ ,  $\Delta X$ , the northward ground magnetic perturbation used to indicate ring current strength, was  $\sim 50\gamma$  near dusk. On the other hand, for a fluctuating  $B_z$  of  $\sim 2\gamma$ ,  $\Delta X$  was only  $\sim 15\gamma$ . They also noted that the partial ring current decay took place within 30 min. of a northward turning of the IMF.

#### 4. DISCUSSION

The measurements at both high and low latitudes provide a consistent picture of the interplanetary-magnetospheric interaction, and because of the strong dependence of convective processes on the IMF, it is a picture which supports the open magnetosphere model. Certainly, the convection field and resulting convective processes are magnified with increasing southward  $B_z$  and solar wind velocity, but the correlation of the convective flow asymmetry with the azimuthal component of the IMF is perhaps even more supportive of the open model since the correlation depends on only one interplanetary parameter,  $B_y$ . Indeed, it is difficult to envisage an energy transfer mechanism which would give rise to such IMF dependences other than one in which the IMF directly couples into the magnetospheric flow. Gonzales and Mozer (1974) showed theoretically that a  $B_y < 0$  ( $B_y > 0$ ) component would cause a clockwise (counterclockwise) rotation of the reconnection line of the dayside magnetopause about the sun-earth line. Such a rotation would shift the magnetospheric flow, and thus enhance the convection field, towards dusk (dawn) in the northern hemisphere, and vice versa in the southern hemisphere. The authors also discussed the potential importance of the  $B_x$  component of the IMF on the position of the reconnection line, although the interdependence between  $B_x$  and  $B_y$  through the IMF spiral structure precludes the possibility of any unambiguous correlative study using  $B_x$  as an additional independent parameter.

The observations also tell us that we are capable of monitoring macroscopic solar wind structures through magnetospheric convective processes. The solar wind sector polarity may be inferred from dawn-dusk asymmetries in the low altitude polar cap electric field, equivalent current systems, or low energy electron fluxes. Increases in the magnitude of convective processes at high and low altitudes and high and low latitudes result from increases in southward  $B_z$  or solar wind velocity, and these interplanetary characteristics are found at the leading edge of solar wind interaction regions. However, the correlation between increases in convection activity and the arrival at earth of interaction regions has not yet been performed.

On the other hand, with the present empirical knowledge we are not able to use convection processes to monitor microscopic solar wind parameters, such as the magnitude of the IMF or  $V_{sw}$ . This is partially due to the difficulties involved in studying the interplanetary-magnetosphere interaction through the convection electric field. Firstly, whereas the convection field is a large-scale structure covering the entire magnetosphere, most of the experimental results involved single point measurements. In addition, most of the single point measurements were indirect; that is, the state of the convection field was inferred from measurements of particles, plasma drifts, and magnetic perturbations. However, these procedures are not necessarily limitations.



Large numbers of point measurements covering a large region of space provide a good statistical base, and the diversity of the measurements serves to corroborate the results.

Most of the measurements were made at low altitudes, and thus some transformation between magnetosphere and ionosphere is required. Such a transformation between the measured or inferred ionospheric electric fields and their counterparts in the equatorial plane is non-trivial because of the existence of electric fields parallel to the geomagnetic field. The lack of knowledge of the precise ionosphere-magnetosphere coupling is not critical since parallel electric fields are important for particle acceleration only in the auroral zone, and even there the associated field-aligned potential drop is only a fraction of that across the entire magnetosphere.

The above difficulties apply to the experimental method; a more important shortcoming, however, is the poor determination of which interplanetary parameters are truly independent ones. The data suggest that  $B_y$  alone is responsible for asymmetries in convective processes across the polar cap and auroral zone, but most studies have treated the magnitude of the magnetospheric response to independent changes in  $V_{sw}$  and  $B_z$  rather than some combination of the two. Consequently, we are at a stage where we can use the convection electric field and convective processes to monitor these individual parameters only qualitatively. We know that increases in  $V_{sw}$  and/or southward  $B_z$  will enhance convection, but a change in convection does not indicate which parameter is responsible. In general,  $B_z$  is treated as the most important variable, and  $V_{sw}$  is regarded with less interest. Perhaps this treatment is justified since there is considerably less variation in  $V_{sw}$  than in  $B_z$ , but certainly other parameters of much importance are the east-west component of the IEF ( $\alpha V_{sw} B_z$ ) and the energy input into the magnetosphere,  $J \cdot \bar{E} (\alpha V_{sw} B_z B_{z'}^{\text{tail}})$ . Consequently, I stress the use of multi-parameter studies to determine the relative importance of the different variables.

Another extremely important aspect of the problem is the ability of the experimenters to present an overall picture of the magnetospheric response. Now this is not a statistical picture, but rather a view of how the entire magnetospheric convection system reacts in a given event. The nature of the coupling of the high to low latitude electric fields can be determined by such a study. Currently, only theoretical work has been done on this important topic (Jaggi and Wolf, 1973; Southwood, 1977).

The IMS program provides a means to follow up on some of the suggestions. The expanded ground magnetic observatory chain gives us the opportunity to monitor convective processes simultaneously over a wide range of latitude and longitude. Of great importance in obtaining an overall magnetospheric picture would be the correlation of simultaneous multi-latitude radar measurements of ionospheric drifts. There will be no dearth of data from space, for experiments aboard GEOS 1 and 2 and ISEE 1 and 2 -- the combination of satellites covering a wide range in L -- are suited to study magnetospheric convective processes. Electric field probes aboard these satellites are sensitive enough to measure the electric field in the equatorial magnetosphere. In particular, we would be able to determine whether the  $B_y$  associated asymmetry observed at low altitudes over the polar cap is observed in space. While measurements are being made on or near the earth, ISEE-3 will be gathering information about interplanetary conditions. We therefore have the capability to learn much more about the convection electric field and associated convective processes, and this knowledge will inevitably lead to a better understanding of the interplanetary-magnetospheric interaction. This

will bring us one step closer to making magnetospheric convection a quantitative diagnostic of interplanetary conditions.

#### ACKNOWLEDGMENTS

I gratefully acknowledge comments and suggestions by R.J. Walker, J.A. Slavin, M.G. Kivelson, and D.J. Southwood. This work has been supported in part by NASA grants numbers NSG 7295 and NGL 05-007-004, and by NSF grant DES 74-23464.

## REFERENCES

- Axford, W.I. and C.O. Hines (1961): A unifying theory of high-latitude geophysical phenomena and geomagnetic storms. Can. J. Phys., 39:1433.
- Bahnsen, A. and N. D'Angelo (1976): Solar wind electric field modulation in the interplanetary sector structure. J. Geophys. Res., 81:683.
- Blanc, M. (1978): Midlatitude convection electric fields and their relation to ring current development. Geophys. Res. Lett., 5:203.
- Burton, R.K., R.L. McPherron, and C.T. Russell (1975): An empirical relationship between interplanetary conditions and  $D_{st}$ . J. Geophys. Res., 80:4204.
- Clauer, C.R. and R.L. McPherron (1978): Predicting partial ring current development, unpublished manuscript, U.C.L.A.
- Coleman, P.J., Jr. (1971): A model of the geomagnetic cavity. Radio Science, 6:321.
- D'Angelo, H. and J. Olesen (1975): On the relation of the "slant E condition" in polar cap ionograms to the solar wind sector structure. J. Geophys. Res., 80:2866.
- D'Angelo, N., M. Møhl Madsen, and I.B. Iversen (1976): Polar cap ionospheric electric field modulation by the solar wind sector structure. J. Geophys. Res., 81:2417.
- Dungey, J.W. (1961): Interplanetary magnetic field and auroral zones. Phys. Rev. Lett., 6:47.
- Fairfield, D.H. (1977): Electric and magnetic fields in the high-latitude magnetosphere. Rev. Geophys. and Space Phys., 15:285.
- Friis-Christensen, E. and J. Wiljhelm (1975): Polar cap currents for different direction of the magnetic field in the Y-Z plane. J. Geophys. Res., 80:1248.
- Gonzales, W.D. and F.S. Mozer (1974): A quantitative model for the potential resulting from reconnection with an arbitrary interplanetary magnetic field. J. Geophys. Res., 79:4186.
- Heppner, J.P. (1972a): Electric field variations during substorms: OGO-6 measurements. Planet. and Space Sci., 20:1475.
- Heppner, J.P. (1972b): Polar cap electric field distributions related to the interplanetary magnetic field direction. J. Geophys. Res., 77:4877.
- Horwitz, J.L., J.R. Doupnik, and P.M. Banks (1978): Chatanika radar observations of the latitudinal distributions of auroral zone electric fields, conductivities, and currents. J. Geophys. Res., 83:1463.



- Jaggi, R.K. and R.A. Wolf (1973): Self-consistent calculation of the motion of a sheet of ions in the magnetosphere. J. Geophys. Res., 78:2852.
- Kivelson, M.G. (1976): Magnetospheric electric fields and their variation with geomagnetic activity. Rev. Geophys. and Space Phys., 14:189.
- Lassen, K. and C. Danielson (1978): Quiet time pattern of auroral arcs for different directions of the interplanetary magnetic field in the Y-Z plane. J. Geophys. Res., 83:5277.
- Maezawa, K. (1976): Magnetospheric convection induced by positive and negative Z components of the interplanetary magnetic field: Quantitative analysis using polar cap magnetic records. J. Geophys. Res., 81:2289.
- Meng, C.I., S.-I. Akasofu, and K.A. Anderson (1977): Dawn-dusk gradient of the precipitation of low-energy electrons over the polar cap and its relation to the interplanetary magnetic field. J. Geophys. Res., 82:5271.
- Mozer, F.S., W.D. Gonzales, F. Bogott, M.C. Kelley, and S. Schutz (1974): High-latitude electric fields and the three dimensional interaction between the interplanetary and terrestrial magnetic fields. J. Geophys. Res., 79:56.
- Mozer, F.S. and P. Lucht (1974): The average auroral zone electric field. J. Geophys. Res., 79:1001.
- Nishida, A. (1975): Interplanetary field effect on the magnetosphere. Space Sci. Rev., 17:353.
- Parks, G.K. and R. Pellat (1972): Correlation of interplanetary  $B_z$  field fluctuations and trapped particle redistribution. J. Geophys. Res., 77:266.
- Paulikas, G.A. and J.B. Blake (1976): Modulation of trapped energetic electrons at 6.6 Re by the direction of the interplanetary magnetic field. Geophys. Res. Lett., 3:277.
- Rosenberg, R.L. and P.J. Coleman, Jr. (1978): Solar cycle-dependent north-south field configurations observed in solar wind interaction regions, U.C.L.A. Institute of Geophysics and Planetary Physics Publ. No. 1804.
- Siscoe, G. and N. Crooker (1974): A theoretical relation between  $D_{\text{st}}$  and the solar wind merging electric field. Geophys. Res. Lett., 1:17.<sup>st</sup>
- Southwood, D.J. (1977): The role of hot plasma in magnetospheric convection. J. Geophys. Res., 82:5512.
- Stern, D.P. (1977): Large-scale electric fields in the earth's magnetosphere. Rev. Geophys. and Space Phys., 15:156.
- Vasyliunas, V.M. (1968): A crude estimate of the relation between the solar wind speed and the magnetospheric electric field. J. Geophys. Res., 73:2529.

## REVIEW OF SELECTED GEOMAGNETIC ACTIVITY INDICES

Joe H. Allen  
World Data Center-A for Geomagnetism  
Boulder, Colorado, U. S. A.

and

J. Feynman  
Physics Department, Boston College  
Boston, Massachusetts, U. S. A.

### INTRODUCTION

For almost as long as scientists have been interested in the causes and effects of geomagnetic variations as recorded at magnetic observatories, there have been attempts to derive summary magnetic activity indices. These have ranged from very simple but subjective classifications of magnetograms from single observatories ("quiet", "moderately disturbed", or "disturbed") to exceedingly complex and expensive-to-derive measures of the global range of potential associated with the equivalent currents which could have produced the variations monitored at a large array of recording sites.

In some cases these indices are substitutes, often imperfectly understood, for direct measures of physical quantities which could not then have been obtained. For example, the AE index (or its components AU and AL) may be related to the energy coupled from the interplanetary medium, via the magnetosphere, into the upper atmosphere and lower altitudes of the auroral zone. In principle, it now should be possible to measure this energy input directly using one or more well-located satellites in the solar wind. However, indices may also be summary measures which graphically and quantitatively combine a complex set of phenomena so that these may be taken into account in comparisons with other types of data. Besides summarizing large masses of data, these indices may also serve to eliminate some components of response (e.g. SQ variations removed in derivation of K indices) and focus attention on the remainder. Also, they may make possible the extraction of statistical behavior patterns which would be obscured by the complexity of the mass of original data.

At this time, our Working Group on Interplanetary-Magnetosphere Interactions is interested in the possibility, now under active consideration, of directly monitoring interplanetary conditions in the solar wind and thereby making possible real-time predictive applications, possibly removing the need for some indices. Also, there is interest in the improvement or better understanding of the physical significance of currently derived indices and in the need for indices derived from observatory records in regions not now represented.

It should be noted that in practical terms no global magnetic activity index is available with sufficient timeliness so as to be useful for short-term predictive applications. Also, even if an index such as AE were available in real time (or some directly measured quantity from a satellite in the solar wind) still this might not be particularly useful for real-time applications because of the large degree of dynamic variability in the location of substorm and magnetic storm phenomena. There are, however, three senses of "predictive" use of magnetic activity indices: (i) predictive of conditions that existed at some past time; (ii) predictive of conditions at the present when indices are derived in real time; and (iii) predictive of conditions that will exist in the future (forecasting). Examples of each type of predictive index application are found in the current literature, especially in the papers submitted for the International Solar-Terrestrial Predictions Workshop. The indices which we have chosen to discuss below are, in general, most often used in type-(i), "historic" predictions (e.g. magnetospheric models for different levels of magnetic activity, historical studies of particular disturbances, statistical studies of types of activity, correlative applications, etc.) and type-(iii), future predictions (e.g. forecasts of the frequency of occurrence of disturbances of a particular amplitude, forecast of maximum sunspot number as function of minimum annual magnetic activity level, etc.).

## BACKGROUND

The International Association of Geomagnetism and Aeronomy (IAGA) has a Division V: Observatories, Instruments, Indices, and Data which takes responsibility for the international adoption of magnetic activity indices and encourages their derivation, archiving, and distribution at several institutions. J. V. Lincoln, Chairman of IAGA Working Group V-6: Indices, has published a compilation of some 50 past, present, and proposed geophysical indices (1977). This comprehensive list includes the magnetic activity indices we have chosen to discuss here and many other less well-known indices. Brief descriptions of the major magnetic activity indices, historical information and references are given in the IAGA Bulletin 32 annual series (continuing the IAGA Bulletin No. 12 series). These Bulletins are edited by D. van Sabben, Director of IAGA's International Service of Geomagnetic Indices (ISGI) at De Bilt, The Netherlands (1978).

We have chosen to discuss only the principal magnetic activity indices currently in use. We comment on their method of derivation, possible problems, degree of redundancy and future possibilities. The indices covered here are grouped by region (auroral zone, mid-latitude, low-latitude) and include: the Auroral Electrojet Indices (AE, AU, AL, and AO); the K-type indices (Kp, Kn, Ks, and Km and their equivalent amplitude counterparts); and Dst. A map showing the networks of stations used for deriving these indices is shown in Figure 1 and the stations are listed in Table 1.



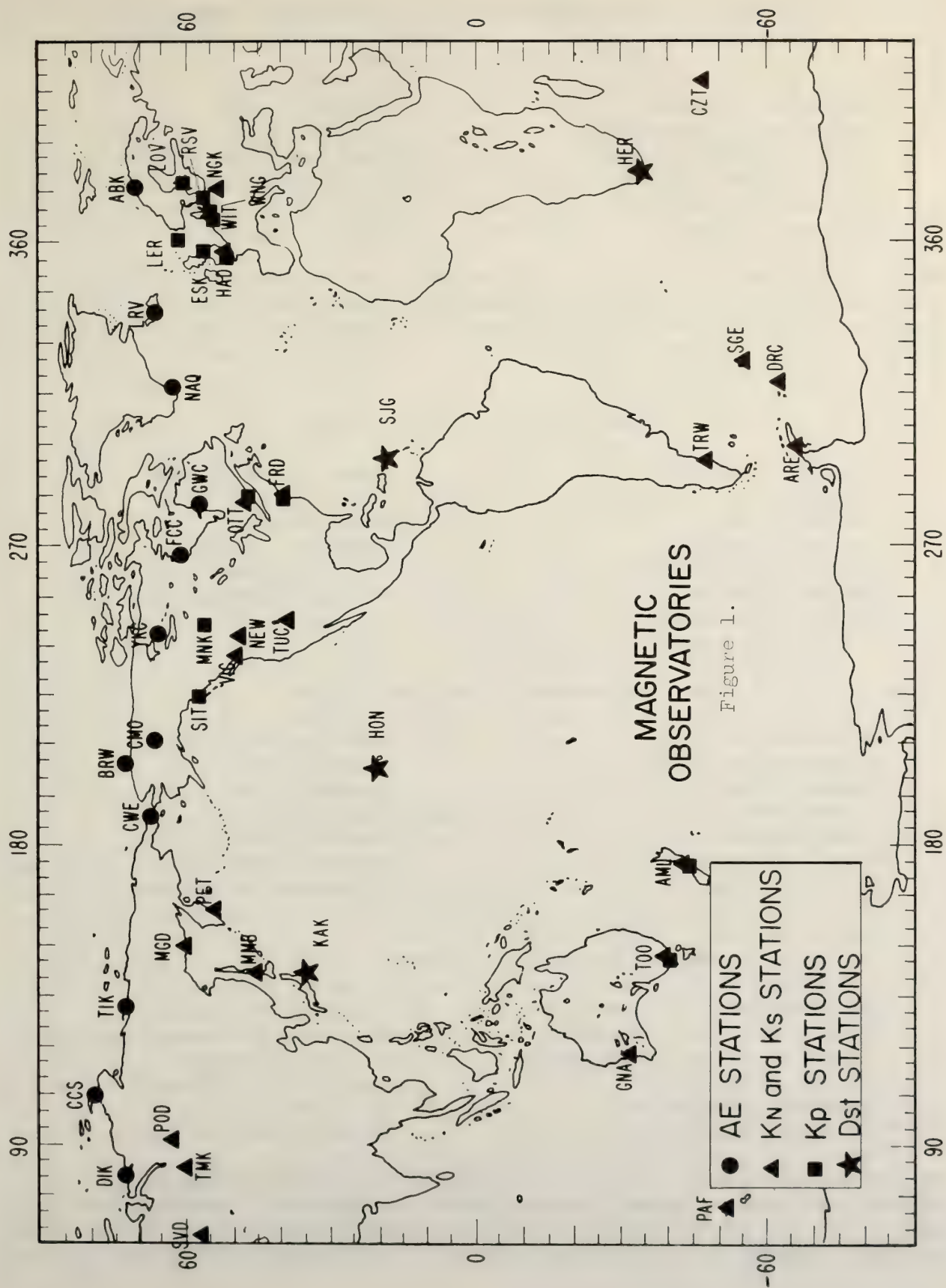


Table 1: Station networks used in index derivations.

AE: -- up to 12 sites --

Leirvogur, Iceland	Barrow, Alaska
Narssarssuaq, Greenland	Cape Wellen (Uelen), USSR
Great Whale River, Canada	Tixi Bay (Tiksi), USSR
Fort Churchill, Canada	Cape Chelyuskin, USSR
Yellowknife, Canada	Dixon Island, USSR
College, Alaska	Abisko, Sweden

Up to 12 stations have been used for the WDC-A derived AE indices. For some times not all these stations were in operation, e.g. Sodankyla, Finland, has been substituted for Abisko; and Yellowknife, Canada, only began operation in 1975.

Kp, ap, and Ap: -- 13 sites --

Lerwick, Shetlands (UK)	Amberly, New Zealand
Eskdalemuir, Scotland (UK)	Toolangi, Australia
Hartland, UK	Lovo, Sweden
Ottawa, Canada	Rude Skov, Denmark
Fredericksburg, USA	Wingst, Germany
Meanook, Canada	Witteveen, Netherlands
Sitka, Alaska (USA)	

Kn, an, and An: -- up to 14 sites --

Hartland, UK	Petropavlosk, USSR
Fredericksburg, USA	Magadan, USSR
Ottawa, Canada	Tunguska, USSR
Tucson, USA	Tomsk, USSR
Newport, USA	Sverdlosk, USSR
Victoria, Canada	Niemegk, Germany
Memabetsu, Japan	Witteveen, Netherlands

Ks, as, and As: -- up to 10 sites --

South Georgia, S. Atlantic	Gnangara, Australia
Orcadas del Sur, S. Atlantic	Kerguelen (Port-aux-Francais), Indian Ocean
Argentine Island, Antarctica	
Trelew, Argentina	Crozet (Port Alfred), Indian Ocean
Amberley, New Zealand	
Toolangi, Australia	Hermanus, South Africa

Indices Km, am, and Am are derived from the corresponding Northern and Southern hemisphere indices from the sites listed above. Depending upon the year, different station groupings have been used to derive these indices for the years 1959-63 (see IAGA Bulletin No. 39), 1964-67 (see INDICES Kn, Ks et

KM, P-N Mayaud, CNRS, 1968), and for years 1968-76 (see IAGA Bulletin No. 32 series (a) through (g) and continuing).

aa: -- 2 sites --

Greenwich, UK, and Melbourne, Australia, from 1868 through 1919.

Greenwich, UK, and Toolangi, Australia, from 1920 through 1925.

Abinger, UK, and Toolangi, Australia, from 1926 through 1956.

Hartland, UK, and Toolangi, Australia, from 1957 to present.

Dst: -- 4 sites --

San Juan, Puerto Rico (USA)

Honolulu, Hawaii (USA)

Kakioka, Japan

Hermanus, South Africa

Dst was derived from a larger station array for the IGY, but the continuing indices are obtained from the 4-station network given here.

#### AURORAL ZONE INDEX

The Auroral Electrojet (AE) index and its associated indices (AU, AL, and A0) are practically the only magnetic activity indices in wide-spread use for auroral zone studies. AE was defined by Davis & Sugiura in 1966. Hourly AE indices were derived at the University of Alaska by Davis and others beginning with IGY (July 1957) and continuing through 1964. At NASA's Goddard Space Flight Center, Sugiura and Fairfield derived 2.5-minute AE indices for part of 1964 and 1965, using data from a relatively complete array of stations. Both the University of Alaska and NASA-derived AE indices used data from a mixture of northern and southern hemisphere stations, irregularly distributed in longitude. For the years 1966 through 1974, J. Allen and staff of World Data Center-A for Solar-Terrestrial Physics (WDC-A for STP) derived 2.5-minute AE indices from as well distributed as possible an array of northern hemisphere auroral zone observatories. Hourly average values and daily graphs of the 2.5-minute index variations were published in WDC-A's series of UAG Reports. Early conclusions from the systematic derivation of AE and the related indices from 1966 through 1971 are reported by Allen & Kroehl (1975). For 1975, derivation of 1-minute AE indices was started and publication of the first UAG Reports for this interval has just begun. Also, preliminary 1-min AE indices from a reduced station network have been published for the first 4 months of the International Magnetospheric Study (Jan - April 1976).

If H-component magnetograms from an array of auroral zone observatories that cover all longitude sectors could be digitized and reproduced to common time and amplitude scales as variations superposed relative to the same quiet-time baseline level, the amplitude at each moment of the upper envelope of traces would define AU and the amplitude of the lower envelope would define AL. The range between upper and lower envelopes is AE (i.e.  $AE = AU - AL$ ) and the mid-trace is A0 (i.e.  $A0 = 0.5(AU + AL)$ ). In practice



this means that for each station used, digital values must be obtained of the H (or X) component variation at the time interval for which AE is desired. A quiet time level must be determined for each site (it will change with time) and subtracted from the scaled values, leaving only the perturbations at each site. For each index time interval, these perturbation values are compared, and the largest (most positive) is AU, the lowest (most negative) is AL for that time.

In one sense AU and AL are local indices because they are only the amplitude of H-component perturbations at two sites. However, these local extremes are determined by comparison with a globe-circling array of data, and in this sense AE is a global auroral zone index. We know these local extreme variations show systematic patterns which define global, statistical features of the current patterns associated with substorms. The extreme positive H variations which give AU are caused by the eastward electrojet and may arise from a station located at local magnetic times (longitudes) from 1200 to 2200 LGT. Most often AU will come from a site near 1800 LGT with the most frequent occurrence of AL from a site near 0300 LGT.

Objections to the AE indices usually are based on the inadequate (non-optimal) spatial distribution of stations whose records are used. During very quiet times the auroral oval contracts poleward and substorms during these times may have magnetic variations that are barely observed or not seen at the latitudes of the AE network (62.5 to 71.6 deg corrected geomagnetic latitude). During large magnetic storms the auroral oval track of the electrojets may expand equatorward of the network. Perhaps equally important is the inherent limitation that a small substorm may have electrojet current systems of limited longitudinal extent but located exactly above AE observatories so that their magnetic contribution to the indices is recorded with optimal efficiency. A larger substorm may have widespread electrojet current systems but be located so that the recorded magnetic perturbations are no larger than for the smaller event. Other problems are that a change in AU or AL (and hence in AE) can result either from an increase in the current flowing in the electrojets or in the location of the electrojets relative to the stations. Also, the auroral zone perturbations monitored in these indices may combine effects of convection currents, tail currents and other magnetospheric currents with the electrojet effects. Seasonal effects may arise from use of data from only one hemisphere (they are certainly seen in the amplitude of AU). And, it may be that results from deriving AE using only the H-component perturbations are not as meaningful as could be obtained from a more appropriate combination of H and D-variations ( $X_m$ ).

In spite of the "problems" outlined here, it remains a fact that AE (and the others) are the most satisfactory auroral zone magnetic activity indices available. Data from new stations can begin to be incorporated as these are created to fill the present gaps (Yellowknife data were first used for 1975 AE(12)). As present analog recording equipment is replaced with digital magnetometers, the work to process the records will be greatly reduced (about 2/3 of the effort is in magnetogram digitizing) and improvements in index timeliness will result. In principle it would be possible to use digital magnetic variations recorded at some 5 or 6 ground sites

and relayed via geostationary satellites to a central collection point to compute a near-real time AE index (similar to the dedicated N. American IMS magnetometer network relayed via SELDADS). Such an index might have useful short-term forecasting applications but probably their full utility could only be realized in conjunction with real-time monitoring of the solar wind from satellites upstream of the earth.

## MID-LATITUDE INDICES

The Kp, ap, aa, and am indices are measures of midlatitude geomagnetic activity. As discussed below they are approximate. However, it should be kept in mind that except for the original records no other index measures the activity in the same frequency range in midlatitudes. A complete understanding of geomagnetic activity must include midlatitudes although there is no role for these indices in short-term forecasting. However, activity at these latitudes is of primary importance, from the viewpoint of human needs, because the midlatitudes are the most densely populated. In the discussion following most of the emphasis is on the difficulties with these indices, however, these should be considered in the perspective of the importance of the indices.

The Kp, ap, aa, and am indices are closely related and are all representations of the level of geomagnetic variability at midlatitudes in the frequency range of about 1/2 hour to several hours. Since Kp and ap were devised in 1939, they are not designed to measure certain specific processes we now envision as basic magnetospheric dynamics, such as the enhancement of the ring current or substorms. During the past ten years, substorms and the ring current have been studied extensively and relatively little attention has been paid to midlatitude disturbances. In fact, the prevailing view has been that midlatitude variations are nothing more than a pale reflection of high latitude events. However, this is apparently not the case. As discussed below, midlatitude and high latitude variations are related to the interplanetary medium in significantly different ways so that it would seem that at least a component of midlatitude activity involves a process distinct from that causing the high latitude and ring current activity.

All of the midlatitude indices are derived from an index K, measured at each station. The meaning and accuracy of the Kp, ap, aa, and am are all dependent on the individual K indices. A review of the process of arriving at K was made by Mayaud (1967a). The index was originally defined as being the range of the most agitated of the three components of the field measured at the station in a 3-hour period. However, it was found that the z-component was too sensitive to the induced ground currents and so only D and H have been used since 1964. We are not aware of studies to evaluate the contribution of earth currents in these indices before and after 1964. The K index may have one of ten values, 0 to 9. Before estimating K, the daily variation Sq and the "post perturbation effect" Dst are both removed from the data. Unfortunately, since the daily variation changes form and amplitude from day to day, there is no objective criterion that can be used to remove it. After Kp is determined, the average quiet day variation is



often determined for stations as the change in the mean hourly values in local solar time for the five international quiet days of the month. Mayaud recommends that each observer use individual judgment in estimating the daily variation for any particular day. The observer must also use judgment in removing the Dst; also, Solar Flare Effects (sfe) are omitted in scaling K.

The scaled K indices are then converted to Ks indices (K - standardized) which takes account of diurnal variations in the level of activity and seasonal effects. These conversion tables have been criticized (Mayaud and Svalgaard, 1978) because the period used to remove the daily variation did not take proper account of the variability of the daily variation.

These Ks indices are, therefore, approximate. The effects of the approximate nature of the Ks will depend on how they are combined to form Kp, ap, aa, and am.

The Kp is the arithmetic mean of Ks from 13 observatories between geomagnetic latitudes of 46 and 63 degrees; eleven observatories in northern latitudes and two in southern latitudes. The averaging will decrease the effects of individual judgment in estimating K. The predominance of northern stations is discomforting, but has not been shown to lead to any serious difficulties.

The ap is derived from Kp by using a conversion table in which the relationship is "quasi-logarithmic," as shown by the crosses in Figure 2. The straight line has been added to indicate the deviations from a logarithmic relation. Note that during quiet times the deviations are large. The ap index is considered to be an estimate of the range (in units of 2 gamma) of magnetic disturbance that would be seen at about 40 degrees of ap for the UT day. Another daily index,  $\Sigma Kp$ , the sum of the eight Kp values in the UT day, is often used; although, because of the quasi-logarithmic nature of Kp, this summation has little physical meaning.

Some of the difficulties of Kp and ap have been avoided in the development of a relatively new index am (Mayaud, 1967b). The am is derived from a Km index, which is like Kp except that the distribution of stations has been much improved. In fact, two sets of K-type indices are derived: Kn, an, and An from a northern hemisphere network of up to 14 sites and Ks, as, and As from a southern hemisphere network of up to 10 sites. The index Km ('m' is for 'mondial', i.e. global) is an average of the indices Kn and Ks for each 3-hour interval of the UT day. The am is derived from Km using a table (Mayaud, 1968) represented graphically in Figure 2, where a straight line is again added to indicate deviations from logarithmacy for small values. The units of am are 1 gamma. Km = 9 for all values of am greater than 600 gammas. The conversion of the data from the individual stations to make up Km is much more simple than that needed for Kp, so less information is lost in constructing Km. For example, it retains a Universal Time variation that has been studied by Svalgaard (1975).

The aa index is derived from two nearly antipodal stations (Greenwich and Melbourne or Hartland and Toolangi) and is an approximation to am (Mayaud, 1971). Its major advantage is that Mayaud has produced a set of



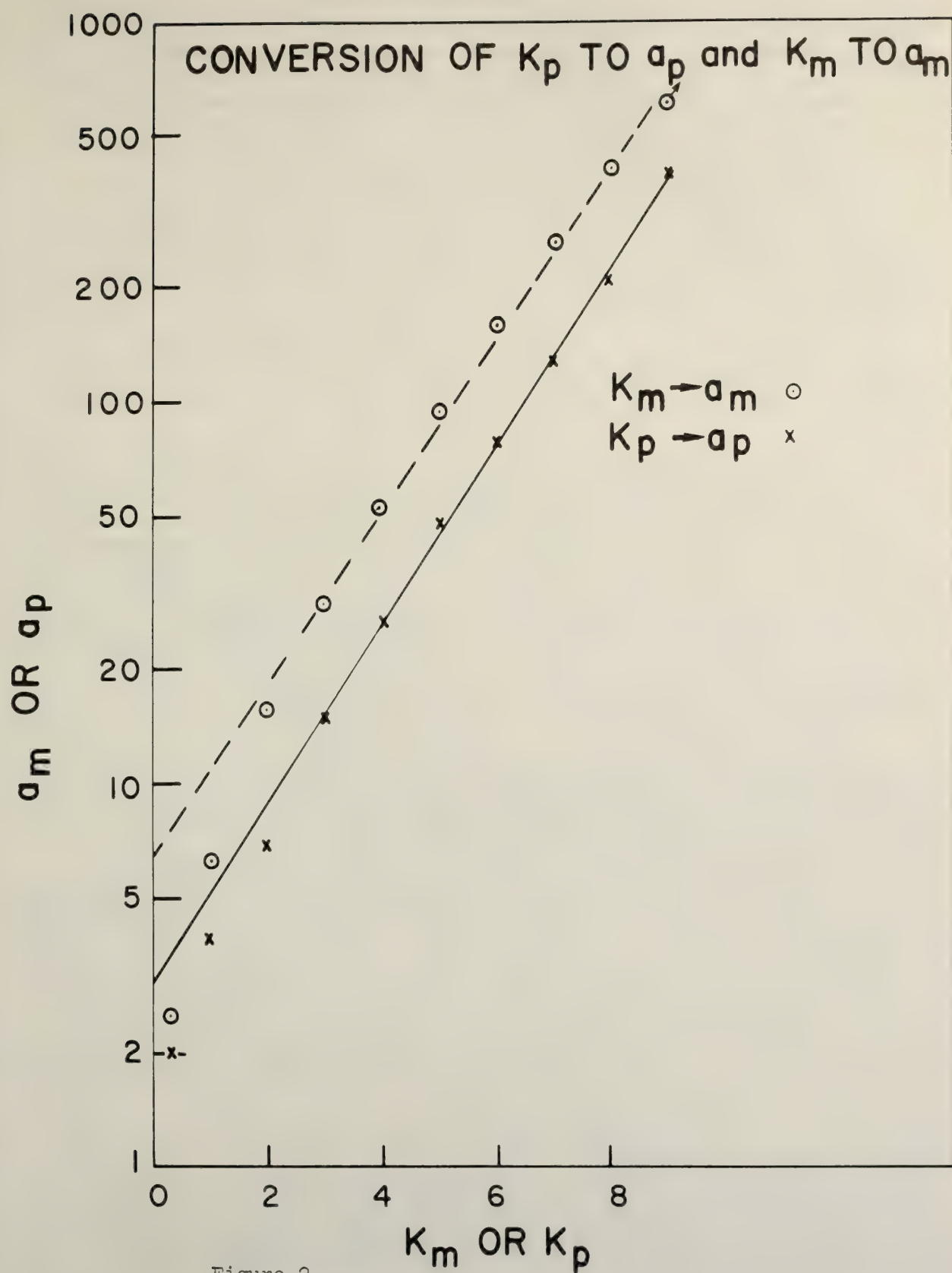


Figure 2.

commensurate indices reaching back to 1868. Although aa is given as a 3-hour index, no individual aa should be considered accurate, and it should only be used in statistical studies where the averaging will decrease the inaccuracies. However, Allen and Kamei have just reported at this Workshop that for the years 1966 through 1974 the aa indices have a correlation coefficient with the corresponding am (or ap) indices of  $r=0.94$ , and that this correlation is relatively stable for all 3-hour intervals of the UT day. Rostoker (1972) in his review of geomagnetic indices has also stressed that Kp and ap should only be used statistically. This may be even more true for aa. However, since the aa series provides a measure of geomagnetic activity which is commensurate over more than 100 years and because it may be obtained relatively quickly from data from only two observatories, it is invaluable for many studies.

Three aa-type daily indices are published by WDC-A. They are averages of 8 3-hourly aa values and are good approximations to the daily average am. The values for the northern and southern hemisphere are given separately.

The frequency distribution of the 3-hour K indices used for aa has been studied by Bubenik and Fraser-Smith (1977). They find the distributions are not the same. For example, the value  $K=1$  is more probable in the south (at Toolangi) than in the north (at Hartland) whereas the  $K=3$  or 4 are more probable in the north. The report of Allen & Kamei also indicated that the northern and southern station inputs to aa are not equal throughout the UT day. The aa index is distinctly more a function of the northern variations during the six hours before and the three hours following geomagnetic midnight at Hartland. Likewise, aa is more a function of variations from the southern hemisphere during the 6 hours before and 3 hours after geomagnetic midnight at Toolangi.

The problems of predicting these indices in real time are being reviewed by Heckman for this workshop and are not discussed here. There have been many studies of the relationship between the indices and parameters in the interplanetary medium and the reference list cited here is incomplete. Recently work on relationships between the solar wind and geomagnetic activity has concentrated on high latitude disturbances and the ring current. However, earlier studies had shown that the midlatitude indices were correlated with various parameters in the solar wind (Wilcox et al., 1967). It was also shown early that the southward field  $B_s$  caused midlatitude activity and that there was an additional dependence on the magnitude of the interplanetary field  $B$  (Hishberg and Colburn, 1969). This dependence of midlatitude activity on  $B$  is extremely important because Maezawa (1978) has recently shown that there is a marked difference between the response of  $A_m$  and the high latitude index  $AL$  to  $B$ . In particular, he shows that  $A_m/AL$  rises steeply with  $B$  when the interplanetary field is northward. Maezawa suggests that  $A_m$  contains the effects of at least two distinct modes of interaction with the solar wind. Various studies using data averaged over time intervals ranging from several hours to six months find correlations between midlatitude activity and  $vB_s$  or  $v^2B_s$  of about 0.8 or 0.9 (c.f. Garrett et al., 1974 and references therein; Crooker et al., 1977) where  $v$  is the solar wind speed. The midlatitude activity is also

correlated with  $\sigma$ , the variance of the interplanetary field (Ballif et al., 1967; Garrett et al., 1974) but considerable controversy still exists as to whether or not this effect is separate from the dependence on the southward field (Garrett et al., 1974; Maezawa, 1978). Comparison of the results of studies is complicated by the many different definitions used for determining the southward field and the variance. The use of interplanetary information in real time prediction is, of course, badly hampered by the lack of real time monitoring data from satellites. Perhaps the best method for "predicting" the midlatitude indices from interplanetary parameters available now would be to take the measured values of the three hour average of the magnitude and velocity  $B$ ,  $v$ , and  $\sin \theta$  (where  $\theta = B_s/B$ ) and the root mean square of the three hour variance  $\sigma$  and consult Maezawa's (1978) figures. While this procedure would not allow a real time prediction, it would test our empirical knowledge of the relationships between midlatitude variations and the interplanetary medium.

### LOW LATITUDE INDEX

The most used and consistently derived index of magnetic activity recorded at low latitudes is Dst which principally represents the axially symmetrical variation in the global H-field due to ring current variations. Dst indices have been derived for studies of special events over many years; however, the principal source of these indices since the IGY has been M. Sugiura. His paper in Annals of the IGY gives references to earlier Dst derivations and discusses different techniques of deriving Dst. In general, long-term secular change in the H-component is removed from low latitude (but not equatorial) station variations (see Table 1 for current network). Fourier analysis of the average disturbance field at the sites gives a zero-order component which is the globally symmetrical disturbance Dst. Normalization by the appropriate function of network latitude gives the "equatorial Dst". If the low latitude disturbance variations were superposed to a common reference level and with common time and amplitude scales, the mid-trace (analogous to A0 in auroral latitudes) would represent Dst.

Not much discussion of Dst took place at the Workshop or in papers prepared for this conference. Problems with Dst have been mentioned that arise from its derivation from so small a network of stations (only 4) and from the relatively coarse time interval for which it is derived (1 hour). The correlative work reported by Allen clearly showed that Dst contained information which could not be extracted from functions of the midlatitude or auroral zone indices. However, it also contained a suggestion that Dst may include a systematic component of variation in UT, perhaps due to incomplete removal of some station local time variations.

### SUGGESTIONS FOR OTHER INDICES

From the discussion above it seems apparent that two regions are not represented by systematically derived, generally available indices. These are the polar cap and the equatorial regions. Experimental polar cap indices of Kokubun & Iijima (1975) were discussed briefly by R. McPherron in a



review presentation at the Workshop. We are not aware of any index of equatorial magnetic activity, but such would presumably give a means for evaluating the quiet to disturbed variations in the equatorial electrojet(s). Another distinctive region in the mapping of the magnetosphere to the earth is the polar cusp. Shelementsev et al. have proposed a polar cusp index (PE) which may prove responsive to large substorms before the AE index changes.

All of the magnetic activity indices discussed in some detail above are for periods of 1-min., 2.5-min., 1-hour, 3-hour, or daily. No pulsation bursts (roughly 60 sec. period and quasi-sinusoidal) accompany the onset of substorms, they might offer a basis for a useful pulsations index.

All magnetic indices discussed have been derived from ground-based recording sites. One of the main prerequisites for any of these sourcesites for an index is that the records be available with a high degree of reliability. The timeliness of their availability and the form (digital or analog) also are important factors affecting what can be derived and how quickly. In principle, satellite magnetometers could provide data for magnetic indices. However, there are usually problems with satellite orbits (often not long in the region where the index would be of interest), difficulties with tracking priority when there are competing satellite programs, and generally a lack of agency or experimenter interest in assuming responsibility for a truly "monitoring" mode of operation. At least two possibilities seem particularly interesting, though, in spite of the difficulties; these are a real-time solar wind and IMF monitoring satellite at the First Lagrangian Point (e.g. ISEE-3) and geostationary satellites such as the SMS/GOES series.

Both of these possibilities are already more or less real. Data collections exist from which trial indices could be derived. It remains for someone with interest, ability, time, and computer-power to perform the calculations, obtain the results, and convince the community of the validity and importance of their indices.

## CONCLUSION

Geomagnetic activity indices have clearly an important role in studies of the interaction between solar activity, the interplanetary magnetic field and solar wind, the magnetosphere, ring current, field aligned currents, and ionospheric currents. It seems worth emphasizing that there is no one single all-purpose magnetic activity index which will serve all needs. In general, there may be said to be three types of indices corresponding to the latitude range of the stations whose records are used for their derivation: auroral zone (AE et al.), midlatitude (Kp, ap, Ap, Km, am, Am, aa, ...), and low latitude (Dst). Obviously, none of these indices are perfect nor likely to become so. Each is a compromise between what was desired and what can be derived given the practicalities that exist (availability of stations, cost of data processing, ...). The high degree of correlation that exists between the several midlatitude indices suggests that unnecessary redundancy may exist. Appropriate international bodies should consider this topic to see if work in deriving competing indices could be combined to reduce the overall expense and improve the timeliness of index derivation. We

especially commend the possibility of more applications of the aa index (or a similar index which could be derived from roughly antipodal digital sites connected via satellite relay links to a real-time processing system). Use of digital networks of observatories connected via satellite relay platforms to a centralized, dedicated mini-computer which can process the individual variations records and derive near-real-time indices from the network seems an excellent idea for the future if the need and utility can be demonstrated. The experience with the N. American IMS magnetometer chain data relayed via the SMS/GOES geostationary satellites to the SELDADS and then to archive at WDC-A shows that this technique can work over approximately 1/2 the globe (low latitude chain).

#### ACKNOWLEDGMENTS

The authors gratefully acknowledge helpful material received from G. Rostoker and R. L. McPherron as well as discussions with our associates on the Working Group on Interplanetary-Magnetosphere Interactions. One author, J. Feynman, acknowledges support from the Air Force Geophysical Laboratory - contract F19628-79-C-0031.

#### REFERENCES

- Allen, J. H. and H. W. Kroehl (1975): J. Geophys. Res. Vol 80, pps 3667-3677.
- Ballif, J. R., D. E. Jones, P. L. Coleman, Jr., L. Davis, Jr., and E. J. Smith (1967): Transverse fluctuations in the interplanetary magnetic field: a requisite for geomagnetic variability, J. Geophys. Res. 72,
- Bartels, J, N. H. Heck, and H. F. Johnston (1939): "The three-hour-range index measuring geomagnetic activity", Terr. Mag. 44, 411-54.
- Bubenik, D. M. and A. C. Fraser-Smith (1977): Evidence for strong artificial components in the equivalent linear amplitude geomagnetic indices, J. Geophys. Res., 82, 2875.
- Crooker, N. U., J. Feynman, and J. T. Gosling (1977): On the high correlation between long-term averages of solar wind speed and geomagnetic activity, J. Geophys. Res. 82, 1933.
- Davis, T. N. and M. Sugiura (1966): J. Geophys. Res. Vol. 71, pps 785-801.
- Garrett, H. B., A. J. Dessler, and T. W. Hill (1974): Influence of solar wind variability on geomagnetic activity, J. Geophys. Res. 79, 4603.
- Hirshberg, J. and D. S. Colburn (1969): Interplanetary field and geomagnetic variations - a unified view, Planet. Space Sci., 17, 1183.



- Kokubun, S. and T. Ijima (1975): The sequence of polar magnetic substorms, Planet. Space Sci., Vol. 23, No. 11, pps 1483-1494.
- Lincoln, J. V. (1977): in IAGA News No. 16.
- Maezawa, K. (1978): Dependence of geomagnetic activity on solar wind parameters: a statistical approach, Solar-Terrestrial Environmental Research in Japan, 2.
- Mayaud, P. N. (1967 a): Atlas of Indices K, IAGA Bulletin, 21.
- Mayaud, P. N. (1967 b): Calcul préliminaire d'indices Km, Kn, et Ks ou am, an, et as, mesures de l'activité magnétique a l'échelle mondiale et dans les hémisphères Nord et Sud, Ann. Geophys., 23, 585.
- Mayaud, P. N. (1968): Indices Kn, Ks, et Km, 1964-1967, ed. du Centre National de la Recherche Scientifique, Paris.
- Mayaud, P. N. (1971): Une mesure planétaire d'activité magnétique basée sur deux observatoire antipodaux, Ann. Geophys. 27, 1, p 67-70.
- Mayaud, P. N. and L. Svalgaard (1978): Comment on Evidence for strong artificial components of the equivalent linear amplitude geomagnetic indices by D. M. Bubenik and A. C. Fraser-Smith, J. Geophys. Res., 83, 2723.
- Rostoker, G. (1972): Geomagnetic indices, Rev. Geophys. Space Phys., 10, 935.
- Sugiura, M.: in Annals of the IGY, Vol. 35, Part I, pps 9-45.
- Svalgaard, L. (1975): Recalibration of Bartels' Geomagnetic activity indices Kp and ap to include universal time variations, SUIPR report #649.
- van Sabben, D. (1978): Brief descriptions of the major magnetic activity indices, historical information and references, IAGA Bulletin 32.
- Wilcox, John M., Kenneth H. Schatten, and Norman F. Ness (1967): Influence of interplanetary magnetic field and plasma on geomagnetic activity during quiet-sun conditions, J. Geophys. Res., 72, 19.



# THE PREDICTION OF AE, ap, AND Dst AT TIME LAGS BETWEEN 0 AND 30 HOURS

D.F.Smart, H.B.Garrett, and M.A.Shea  
Air Force Geophysics Laboratory  
Hanscom Air Force Base, Massachusetts, USA 01731

AE, ap, and Dst are correlated with ~ 35,000 hours of interplanetary plasma and magnetic field measurements acquired near the earth (assembled by NASA/NSSDC into a composite data set). Lag times between the indices and solar wind parameters ranged from 0 to 30 hours. Correlations at lags less than 6-hours yield results in agreement with previous studies. At greater lags, the correlation coefficients between the solar wind parameters and AE and ap approach these parameters' autocorrelation (persistence) values. For Dst, the correlation with solar wind parameters is lower than that with AE and ap in the 3 to 4 hour lag range whereas the autocorrelation of Dst is significantly higher over the entire 0 to 30 hour lag range. The implications of these differences between AE, ap, and Dst are discussed in terms of persistence of solar wind structure.

## 1. INTRODUCTION

The prediction of geomagnetic activity based on solar variations has been a goal of magnetospheric physics since the relationship between sunspots and auroral activity was inferred over a century ago. Since the discovery of the solar wind, many attempts have been made to correlate various solar wind parameters with geomagnetic activity (Schatten and Wilcox, 1967; Ballif et al., 1967, 1969; Hirshberg and Colburn, 1969; Arnoldy, 1971; Foster et al., 1971; Garrett, 1973; Garrett et al., 1974; Bobrov, 1973; Burton et al., 1975; to name a few). Although all have met with some success, there has been much disagreement between them as to what are the critical parameters. This may be in part traced to the relatively small data bases that were used. More recent studies have made use of the much larger solar wind data base for the period 1963-1975 (King, 1977). Besides the present paper, recent such studies are those of Crooker et al. (1977), Maezawa (1978), and Garrett et al. (1978). Although limited to 1 hour resolution and the necessity of relying on many satellites, some of which are not well calibrated, the NSSDC data base allows a gross statistical study of the solar wind-geomagnetic activity relationship at time lags of up to months. In this paper we will make use of the NSSDC data base to study the response of the geomagnetic indices AE, ap, and Dst to various solar wind

parameters at time lags of up to 30 hours.

The purpose of this study will be to determine the differences which exist between the responses of AE, ap, and Dst to the solar wind. Specifically, the persistence of each of these parameters as reflected in their autocorrelation coefficients will be compared to their cross-correlation with various solar wind parameters at lags of up to 30 hours. By doing this we plan to link the data-limited, though detailed, correlation studies of Garrett, et al. (1974), with the solar cycle study of Crooker et al. (1977), which used 6 month averages of the Ap index and solar wind speed. Also we hope to uncover the cross-over point between  $\langle B_z \rangle$ -related (see later), and  $V^2$  (or bulk properties) dominated-correlations.

In the first part of the paper we will describe the data base and various inconsistencies that are known to exist. In the second section we will briefly describe the analytic procedure employed and the parameters that will be studied. Subsequent sections are devoted to analysis of the various correlations that are found. The paper will conclude with a brief attempt to explain the observed correlations.

## 2. DATA BASE

The data base consists of approximately 50,000 hours of simultaneous solar wind plasma data and magnetic field data, 1-hour averages of AE and Dst, and 3-hour averages of ap for the time period 1963 to 1975. Approximate time periods and the satellites involved are listed in Table 1. Not only may data for a given day be from more than one satellite, but hourly values may also be from more than one satellite. In this section we will discuss such problems of coverage and how they might affect the results. Also, problems with individual satellites and with the geomagnetic indices will be covered. For a detailed treatment of how the data were normalized, the reader is referred to King, (1977).

### 2.1 Satellite Data

In Table 1, the approximate time periods for data coverage by the satellites used in the study were listed. The original tape formed from these satellites (under the direction of Dr. J. King of the NASA/NSSDC) was written assuming that certain satellites had higher quality data than others so that a preference system was used when overlapping data were found. Although this assured a better overall product, the results are not necessarily consistent as the different satellite had different calibration accuracies.

Table 1. Source Plasma Data Set Characteristics

Spacecraft	P.I. (Institution)	Averaging Time (Hours)	Time Period	Number of Hours	T	N	V	$\phi_v$	$\theta_v$	$\sigma_T$	$\sigma_N$	$\sigma_V$	$\sigma_\phi$	$\sigma_\theta$
Explorer 18	Bridge (MIT)	3	11/27/63 - 2/22/64	1,485		20%	10%							
Merged Vela	Bane (LASL)	3	7/21/64 - 3/18/71	10,273			X							
Vela 3	Bane (LASL)	3	7/26/65 - 11/13/67	5,721	10%	25%	5%	1.6°		X	X	X	X	
Explorer 33	Bridge (MIT)	1	7/6/66 - 9/23/69	5,637	X	X	X	X	X	X	X	X	X	X
Explorer 34	Ogilvie (GSFC)	1	6/3/67 - 12/16/67	2,282	X	10%	3%			X	X	X		
Explorer 35	Bridge (MIT)	1	7/28/67 - 7/3/68	3,642	X	X	X	1.5°	.75°	X	X	X	X	X
OGO 5	Neugebauer (JPL)	1	3/5/68 - 4/29/71	2,564	15%	8%	1%	X	X					
HEOS 1	Bonetti (CNR, Italy)	3	12/11/68 - 4/15/70	3,142		X	X							
Explorer 43	Bane (LASL)	1	3/18/71 - 3/27/73	7,998	X	X	X							
Merged IMP	Bane (LASL)	3	3/18/71 - 12/31/74	8,539	15%	30%	2%							
Explorer 50	Bridge (MIT)	1	12/1/73 - 12/30/75	9,895	X	X	X	X	X	X	X	X	X	X



King carefully normalized the plasma data (see King, 1977) whereas no correction was applied to the magnetic field data. He found the major calibration errors in the magnetic field to be in the magnetic field components measured by Explorers 18 and 35. King (1973) indicates errors of  $-1 \gamma$  and  $-0.7 \gamma$  in  $B_{ZSE}$  for each of these satellites respectively. This error and another minor error involving the inclusion of some magnetosheath data were corrected before data analysis began. It was found in fact that the change in correlation introduced by these changes was less than 2%, well within the error normally associated with the correlation coefficient (see later). Based on this and the tremendous size of the data base (the least number of points used was 30,000) this finding implies that although individual calibration errors may still exist between the satellites (see, however, Diodato *et al.*, 1979), they will have little affect on the final data.

Another possible source of error is introduced by the fact that the satellites do not provide precisely the same data. The best example of this is in the calculation of the standard deviation of the magnetic field. For the IMP and AIMP satellites the root-mean-square standard deviations are given (though even for these satellites differing time scales were employed):

$$\sigma_{IMP} = [\sigma_{BX}^2 + \sigma_{BY}^2 + \sigma_{BZ}^2]^{1/2} \quad (1)$$

Where  $\sigma_{BX}$ ,  $\sigma_{BY}$ , and  $\sigma_{BZ}$  are the standard deviations of individual rectangular coordinates. For HEOS only the standard deviation in the magnitude average ( $\sigma_B$ ) and angle averages ( $\sigma_\theta$  and  $\sigma_\phi$ ) are given. Fortunately the following expression has been shown to be analytically equivalent to equation 1 (King, 1977):

$$\sigma_{HEOS} = [\sigma_B^2 + B^2 - F^2]^{1/2} \quad (2)$$

Where  $B$  is the average field,  $\sigma_B$  its standard deviation, and  $F$  the length of the vector constituted by the average cartesian components. Again considering that HEOS represents less than 10% of the data base, this difference should not greatly affect the results.

## 2.2 Geomagnetic Parameters

Although the largest errors probably exist in the solar wind parameters, large errors also exist in the geomagnetic indices. As many of these are covered in detail in the review by Rostoker (1974), we will only give a brief discussion here of the indices and some of the problems associated with their derivations. It should be remembered that, as Rostoker and others have noted, all indices measure the minimum level of geomagnetic activity rather than

the maximum -- that is, the geomagnetic activity level indicated by a given index is the lowest that the magnetic field will have.

By way of review, the 3-hour ap index is derived by a direct conversion from the Kp index. The Kp index is derived from the 3-hour  $K_s$  indices: indices which are derived from the maximum deviations in 3-hours in  $B_x$ ,  $B_y$ , and  $B_z$  at the earth's surface. Kp is then the average of the 13 stations returning  $K_s$ . Unfortunately, although local time and seasonal effects are supposed to have been corrected, residual effects are still observed. Further, Kp (or ap) is a mid-latitude index and thus is affected by both the ring current and the auroral electrojet. Also, the conversion scale from the semi-logarithmic index Kp to the supposedly linear index ap is somewhat arbitrary.

AE has been defined so that it measures primarily the variations in the auroral electrojet. It is derived from the 2.5 minute averages of the H-component (or X-component) of the magnetic field in the auroral zone corrected for local time variations. The maximum and minimum average values of the field are determined for a given set of stations for each 2.5 minute interval. The sum of the absolute values of the maximum and minimum is found. This envelope is AE. The station coverage for AE in longitude is generally poor making it a strong function of local time. Likewise, the method of correcting each 2.5 minute period for local time variations by subtracting the average of the 5Q days each month is not very good. Recent work also indicates that the auroral zone changes position, moving polarward with decreasing activity - often away from the AE stations. The hourly values were obtained by simply averaging the 2.5 minute values.

Dst is defined so that it measures primarily the variations in the ring current. Roughly, it is calculated from equatorial or low latitude stations in 2.5 minute intervals by subtracting the average  $S_q$  and permanent field from the disturbed magnetic field. It, too, suffers from poor station coverage and local time variations. Again, the 2.5 minute values were averaged to give one-hour values.

Besides local time effects, other problems also exist. There is, for example, a lack of AE values before 1966 and a change in stations contributing to AE. Likewise, as a side effect of this study, inconsistencies between published values of ap were found. Further, ap is a 3-hour index instead of the 1-hour index it is treated as in this paper (we are currently investigating this effect on our results). Even when all these variations are considered, together with the local time problems, it is again unlikely whether any systematic errors are observable as we have used such a large data base which samples over the entire day. (Note: if Figure 1, 2, and 3 are carefully examined, a 24 peak in the AE, ap, and Dst

autocorrelations is evident - the peak is near the error in the data set, though).

### 3. ANALYTIC PROCEDURE

In the preceding section, various errors in the data base were discussed. It was stated that few of these errors would effect the results and, considering the large size of the data base, none would likely be greater than the normal noise level. In this section, we will define the parameters studied, the analysis techniques employed, and attempt to define the meaning of this noise level (i.e., the order of accuracy to which a correlation coefficient can be determined).

#### 3.1 Parameters

As an exhaustive test of solar wind parameters would be impossible, we have chosen a cross-section of the parameters found by others to give high correlations. They are:

1.  $B_Z(SM)$       The Z component of the interplanetary magnetic field measured in the solar magnetospheric coordinate system.
2.  $B_Z(SE)$       The Z component of the interplanetary magnetic field measured in the solar ecliptic coordinate system.
3.  $\langle B_Z(SM) \rangle$       The "southward" component of the interplanetary magnetic field measured in solar magnetospheric coordinates.  
$$\langle B_Z(SM) \rangle = B_Z(SM) \text{ if } B_Z(SM) < 0$$
$$\langle B_Z(SM) \rangle = 0 \text{ if } B_Z(SM) > 0$$
4.  $\langle B_Z(SE) \rangle$       The "southward" component (see 3) of the interplanetary magnetic field measured in solar ecliptic coordinates.
5.  $V$       The solar wind velocity in units of Km/sec
6.  $V \langle B_Z(SM) \rangle$       The product of the solar wind velocity and the "southward" component of the interplanetary magnetic field (in solar magnetospheric coordinates).



7.  $\sigma_B$  The RMS standard deviation of the interplanetary magnetic field vector (Note:  $\sigma_B$ (HEOS) was determined by different procedure).
8.  $V \sigma_B$  The product of the solar wind velocity and the RMS standard deviation of the interplanetary magnetic field vector.
9.  $T$  The solar wind temperature ( $^{\circ}\text{K}$ ).
10.  $\rho$  The solar wind density ( $\text{n/cm}^3$ ).
11.  $\rho V^2$  Proportional to the solar wind kinetic pressure.
12.  $B$  The magnitude of the average interplanetary magnetic field vector (in gammas).
13.  $B^2$  Proportional to the solar wind magnetic pressure.
14.  $V^2 \langle B_Z(\text{SM}) \rangle$  The product of the square of the solar wind velocity and the southward component of the interplanetary magnetic field (measured in solar magnetospheric coordinates).
15.  $VB^2 \sin^4(\theta/2)$  Parameter defined by Perrault and Akasofu(1978).

In all cases the one-hour values on the NASA/NSSDC tape were assumed.

AE, Dst, and ap were all provided by WDC-A, Boulder, Colorado. They were the latest tapes as of January 1979. The Kp data are from the European Space Agency's European Space Operations Center (Lenhart, 1968). As already mentioned discrepancies between this tape and published ap values were detected though the differences are relatively minor and should not impact our study.

### 3.2 Analytic Methods

Regression analysis (or multiple linear correlation analysis) is a fairly sophisticated mathematical tool. It is not possible, therefore, to cover in any detail the process of regression analysis in this report. The reader is referred to Draper and Smith (1966) or Bevington (1969) for a detailed treatment of regression analysis techniques.

The following analytic procedure was followed in our analysis. That is, step by step the individual cross-correlation coefficients were computed. Next the highest correlator was found and removed. The process was iterated until the addition of further variables failed to make any significant improvement in the multiple

regression coefficient. The uncertainty in the regression coefficient  $R$  was assumed to be given by:

$$\sigma_R = \left( \frac{1 - R^2}{N} \right)^{1/2} \quad (3)$$

where  $N$  is the number of points. Typically  $\pm 2 \sigma_R$  or  $\pm 3 \sigma_R$  was assumed as the range of uncertainty. As described, when the addition of a new variable failed to improve  $R$  by  $+3 \sigma_R$ , the regression was terminated.

#### 4. RESULTS

The data base to be studied consists of approximately 45,000 one-hour values of the plasma and magnetic field. Requiring that all plasma and magnetic field parameters be present in an hour ( $T$  is missing in some cases) reduces the total to 35,000. Requiring that AE be present (ap and Dst are complete over the required interval) reduces the total to 30,000. These last 2 data sets of 35,000 and 30,000 points were intercorrelated. The results of this intercorrelation process will be discussed in 2 parts. The first part will be concerned with the detailed cross-correlations obtained while in the second section the results of the multiple regression analysis will be discussed.

##### 4.1 Linear Correlations

The simplest analytic approach to the massive amount of data represented by the NASA/NSSDC tape is to carry out cross-correlations between all of the pairs of parameters listed in section 3.1 and AE, ap, and Dst. This has been done and the results for 0 lag for the solar wind parameters with each other listed. Also, shown are the cross-correlations with AE and ap at one-hour lag (i.e., the solar wind parameters one-hour before the AE and ap indices were correlated with the indices) and Dst at two hours lag. For ap and Dst, it did not matter whether the 30,000, 35,000, or even the 45,000 point set was used as the results were the same to the accuracy of  $R$  (namely  $\pm 3 \sigma_R \sim \pm 2\%$ ).

The cross-correlations in Table 2 are similar for AE and ap to those calculated in Garrett *et al.* (1974) and many previous studies. The Dst correlations are new results as far as we have been able to determine. The results for inter-correlations between parameters are not too surprising although  $T$ , for example, exhibits rather high correlations with  $V$ ,  $\sigma$ , and ap (this may be related to an increase in  $T$  at the shock boundary of high velocity streams).

Table 2. Cross-correlations between various solar wind parameters and between the solar wind parameters and AE, ap, and Dst.

	$B_z(SM)$																		
1. $B_z(SM)$	1.	$B_z(SE)$																	
2. $B_z(SE)$	.90	1.0	$\langle B_z(SM) \rangle$																
3. $\langle B_z(SM) \rangle$	.82	.74	1.	$\langle B_z(SE) \rangle$															
4. $\langle B_z(SE) \rangle$	.73	.82	.88	1.	$v$														
5. $v$	.01	.00	.04	.04	1.	$v \langle B_z(SM) \rangle$													
6. $v \langle B_z(SM) \rangle$	.80	.72	.97	.85	-.09	1.	$\sigma_B$												
7. $\sigma_B$	.00	.01	-.12	-.11	.29	-.17	1.	$v \sigma_B$											
8. $v \sigma_B$	.00	.01	-.09	-.08	.57	-.18	.93	1.	$T$										
9. $T$	.00	.01	-.02	-.01	.71	-.12	.45	.65	1.	$\rho$									
10. $\rho$	-.01	.00	-.14	-.14	-.34	-.10	.26	.10	-.14	1.	$\rho v^2$								
11. $\rho v^2$	-.01	.00	-.15	-.14	.25	-.20	.52	.54	.35	.75	1.	$v^2 \langle B_z(SM) \rangle$							
12. $v^2 \langle B_z(SM) \rangle$	.73	.65	.88	.77	-.21	.97	-.20	-.25	-.22	-.05	-.23	1.	$B^2$						
13. $B^2$	.01	.01	-.33	-.32	.06	-.34	.51	.46	.25	.30	.46	-.33	1.	$v B^2 \sin^4 \theta / 2$					
14. $v B^2 \sin^4 \theta / 2$	-.55	-.43	-.65	-.52	.14	-.70	.38	.39	.27	.16	.35	-.71	.62	1.	$B$				
15. $B$	.00	.01	-.34	-.33	.09	-.35	.56	.50	.29	.31	.47	-.33	.93	.57	1.	$AE_1$			
16. $AE_1$	-.57	-.51	-.62	-.55	.42	-.67	.33	.42	.38	.04	.33	-.68	.28	.52	.34	1.	$ap_1$		
17. $ap_1$	-.37	-.32	-.51	-.45	.43	-.59	.47	.57	.44	.14	.49	-.64	.49	.62	.49	.68	1.	$Dst_2$	
18. $Dst_2$	.38	.30	.46	.39	-.39	.51	-.24	-.34	-.32	.17	-.13	.53	-.32	-.47	.34	-.50	-.60	1.	

The analysis represented by Table 2 has been carried out at lags of from 0 to 29 hours for all the parameters in Table 2 with ap, AE, and Dst. Likewise, the autocorrelations (i.e., a parameter is correlated with itself at a given lag) have also been found. The results for AE, ap, and Dst are plotted in Figures 1, 2, and 3 along with their highest correlations. As suggested by Feynman (private communication) based on data from Garrett *et al.* (1974), AE and ap correlate in a similar manner for some parameters (i.e.,  $V^2 \langle B_z(SM) \rangle$ ) and differently for other parameters ( $V \sigma$ ). As might have been anticipated, Dst behaves in a very different manner than ap and AE (see, for example, Burton *et al.*, 1975; in a future paper we hope to investigate the rate of change at Dst as suggested in that study). The specifics of these differences will be discussed later - particularly the cross-overs near 6-10 hours lag for ap and AE.



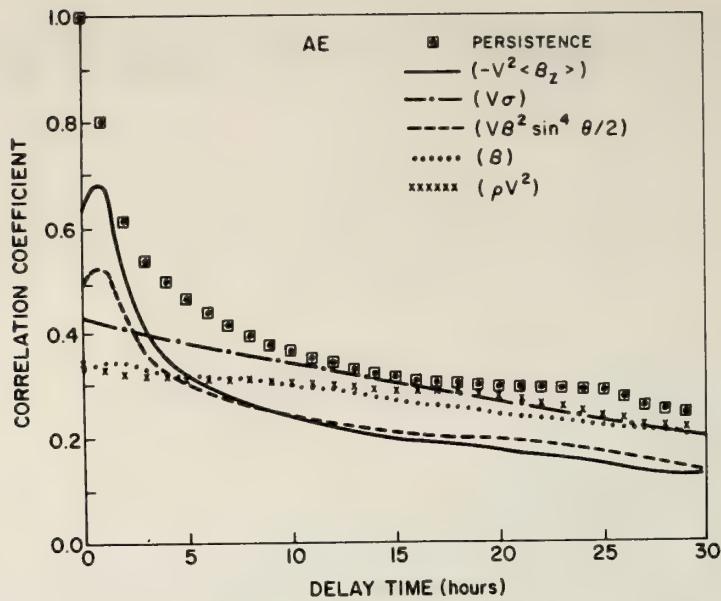


Figure 1. The auto-correlation and cross-correlation coefficients as a function of lag (delay time) for AE with various solar wind parameters.

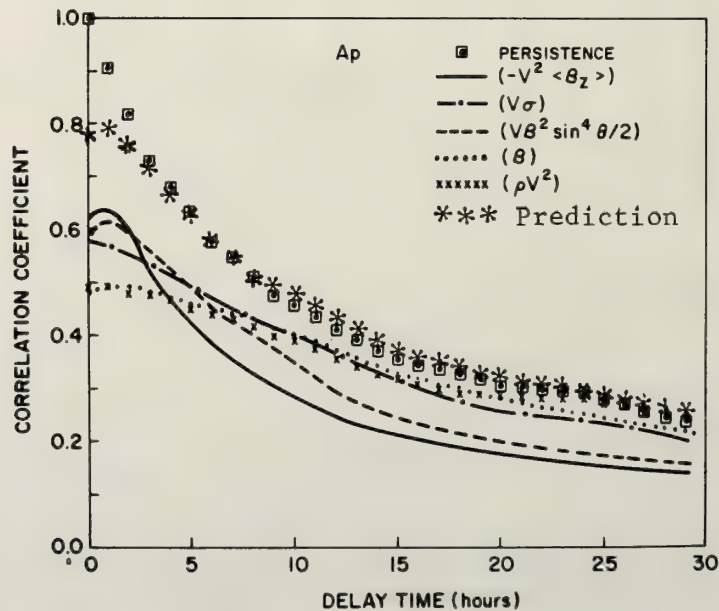


Figure 2. The auto-correlation and cross-correlation coefficients as a function of lag (delay time) for ap with various solar wind parameters. The "\*\*\*" are the cross-correlation coefficients of ap with equation 4.

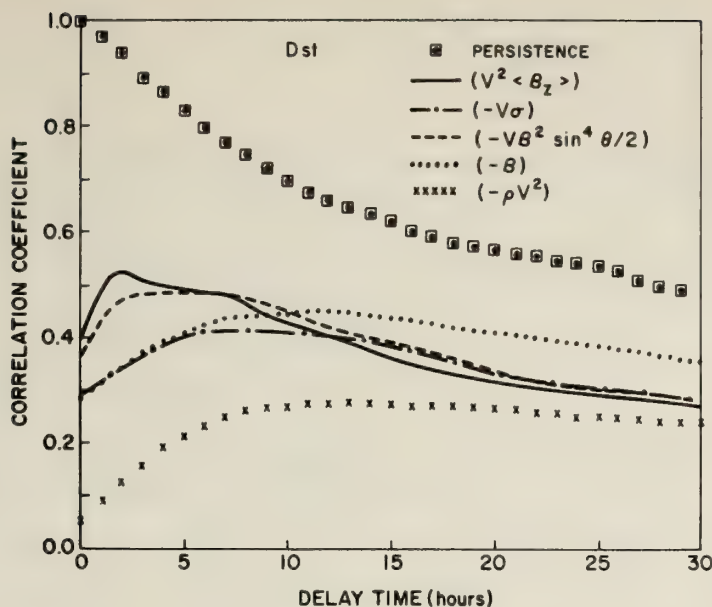


Figure 3. The auto-correlation and cross-correlation coefficients as a function of lag (delay time) for Dst with various solar wind parameters.

## 4.2 Multiple Regression Results

Perhaps more interesting than the cross-correlation coefficients with individual parameters are the results for multiple correlation coefficients. Adding even one more parameter significantly improves our ability to predict ap, AE, and Dst (ap improves from a correlation of 0.63 for  $V^2 B_z$ (SM) to 0.80 by the simple addition of  $V \sigma$ ). Although in principle by employing N variables to describe N data values a perfect fit can always be obtained, generally after the inclusion of only 3 or 4 independent variables the correlation coefficient changes by a lot less than  $\sigma_p$ . Thus it is straight forward to determine the most significant set of independent variables. In this section we will briefly discuss some of the results of our multiple regression.

In Table 3 we have listed the 3 highest correlators at three-hour intervals for AE, ap, and Dst. This table clearly demonstrates the change over from vector-like quantities (i.e.,  $\langle B_z \rangle$  and related parameters such as  $V^2 \langle B_z \rangle$ ) to bulk properties (i.e.,  $\rho V^2$ ,  $V$ , or  $B$ ) with  $V \sigma$  serving as an intermediary correlation. AE and ap are remarkably similar as to what are the best correlators (though the chart correlation values are different) whereas Dst does not correlate at all with  $\rho V^2$ .

Table 3. Highest 3 correlators\*

LAG	AE			ap			Dst		
0	$V^2 B_Z$	V	$\langle B_Z \rangle$	$V^2 \langle B_Z \rangle$	V $\sigma$	$\rho V^2$	$V^2 \langle B_Z \rangle$	V	$\rho$
3	V $\sigma$	V $\langle B_Z \rangle$	V	V $\phi$	V $\sigma$	$V^2 \langle B_Z \rangle$	$V^2 \langle B_Z \rangle$	V	B
6	V $\sigma$	V $\langle B_Z \rangle$	V	V $\sigma$	V $\phi$	$\rho V^2$	$V^2 \langle B_Z \rangle$	V $\sigma$	B
9	V $\sigma$	$V^2 \langle B_Z \rangle$	V	V $\sigma$	B	$V^2 \langle B_Z \rangle$	V $\phi$	V	B
12	V $\sigma$	$\rho V^2$	V	B	V	$\rho V^2$	B	V	V $\langle B_Z \rangle$
15	V $\sigma$	$\rho V^2$	V	B	V	$\rho$	B	V	$\langle B_Z \rangle$
18	V $\sigma$	$\rho V^2$	V	$\rho V^2$	B	V	B	V	$\langle B_Z \rangle$
21	$\rho V^2$	V	B	$\rho V^2$	B	V	B	V	$\langle B_Z \rangle$
24	$\rho V^2$	V	B	$\rho V^2$	B	V	B	V	$\langle B_Z \rangle$
27	$\rho V^2$	V	B	$\rho V^2$	B	$B^2$	B	V	$\langle B_Z \rangle$
30	$\rho V^2$	V	B	$\rho V^2$	B	$B^2$	B	V	$\langle B_Z \rangle$

\* All  $\langle B_Z \rangle$  are solar magnetospheric coordinates  
 $V \phi = VB^2 \sin^4 \theta/2$

In Figure 2 we have plotted the regression curve for the top 5 correlations at 3-hour lag for ap. The formula used was:

$$\begin{aligned} \text{ap}^* = & -4.12 + 1.12 \times 10^{-4} VB^2 \sin^4 \theta/2 + 4.84 \times 10^{-3} \sigma \\ & -1.22 \times 10^{-5} V^2 \langle B_Z \rangle + 3.04 \times 10^{-6} \rho V^2 + .621B \end{aligned}$$

\* $\langle B_Z \rangle$  is in solar magnetospheric coordinates.

The interesting feature of Figure 2 is that for ap the regression prediction approaches the autocorrelation coefficient at about 8-hours lag. Similar results were obtained for AE and Dst but will be presented in a future paper. A possible interpretation of the ap result is given in the next section.



## 5. CONCLUSION

A number of correlations have been carried out in the preceding sections. The results can be summarized as follows:

1) Although many correlations were studied between the solar wind parameters, apparently no marked correlation exists between the magnetic field and plasma parameters whereas strong correlations exist between the parameters in each group (note, however,  $\sigma$ ,  $B$  and  $T$  correlate to some degree).

2) Of the single descriptive solar wind parameters,  $V$  and  $B$  correlate the strongest with geomagnetic activity.

3) Of the combination parameters,  $V^2 \langle B_z \rangle$  consistently correlates the highest at short time lags. At about 6-hours,  $V \sigma$  appears. At greater than 15-hours,  $\rho V^2$  appears for AE and ap as the highest correlator. For Dst,  $B$  becomes the highest correlation ( $\rho V^2$  is very low).

4) The auto-correlation coefficients for ap and AE are equaled (or exceeded) after about 6 to 10 hours by some single correlations and by the multiple regression formulas. Dst always exceeds its correlators.

5) There are similarities but also striking differences between AE and ap and particularly between AE or ap and Dst.

The fact the AE, ap, and Dst do not agree as to their correlations most likely indicates that different geophysical processes are responsible for each. Further, the change in correlators with time has been observed before (Crooker *et al.*, 1977). What is interesting is that in AE and ap (and to some extent in Dst) the fundamental nature of the correlations appears to change between 6 and 10 hours. A clue to the reason for this is seen in the nature of the quantities studied. As mentioned earlier there is a change from vector-like quantities (i.e.,  $\langle B_z \rangle$  and related quantities) to bulk properties of the solar wind ( $\rho V^2$ ,  $B$ ,  $V$ , etc.) with an interim period in which  $V \sigma$  dominates. A possible explanation may be found in Dessler (1967) (see expanded form in Garrett *et al.*, 1974). As discussed there, such a relationship may be intrinsic to the solar wind. That is, if storms are due to more than  $\langle B_z \rangle$ -related variations, then a likely candidate is stream structures, particularly shocks and high velocity streams.  $\langle B_z \rangle$  is a rapidly fluctuating variable ( $\sim 1$  hour to 3 hours) whereas shocks or high velocity streams are many hours to even days in extent.  $V \sigma$  (Garrett *et al.*, 1974) is believed to mark the forward edge of such structures.

In Garrett (1974) it was demonstrated that  $\langle B_z \rangle$  in conjunction with high  $g$  gave higher geomagnetic activity levels. Likewise,  $V\langle B_z \rangle$  and  $V^2\langle B_z \rangle$  have been shown in several studies to give higher correlations than  $\langle B_z \rangle$ . Taking all of this together, a consistent argument can be made that short term variations (on the order of a few hours) are controlled primarily by  $\langle B_z \rangle$ -related parameters. Underlying this relation is the fact that high velocity streams and shocks, which are heralded by plasma turbulence ( $\sigma$ ,  $T$ , etc.), encourage increased activity - thus the appearance of these variables after 6 to 10 hours.

The Dst correlations are not difficult to explain in light of the above. As discussed in Burton et al. (1975) and by Iyemori and Maeda (1979, this conference), Dst has a characteristic time scale or decay time of several hours. That is, the frequency characteristics of Dst are fundamentally different than AE and ap as it responds primarily to the impulsive build up and slow decay of the ring current. Thus it is not unanticipated that Dst would not respond as well to the vectorial quantity  $\langle B_z \rangle$  (though see Burton et al., 1975) at short time scales. On longer time scales, as before, the bulk properties emerge as the major correlators (namely B). As Dst has long time scales inherent in it (Iyemori and Maeda, 1979), its autocorrelation would be higher at longer lags than AE and ap.

To summarize, our results indicate that at short time scales ( $\leq 6$  hours) the auroral electrojet responds to  $\langle B_z \rangle$ -related variations (particularly  $V^2\langle B_z \rangle$ ). At greater time periods, this correlation drops and is replaced by correlations with the bulk properties of the solar wind. The ring current, as measured by Dst, responds on a different time scale. At short time scales, although  $\langle B_z \rangle$ -related quantities dominate (crossing at about 6-10 hours), the correlation with solar wind parameters is weaker than found with AE and ap (the correlation with  $\rho V^2$  is particularly low). At longer lags, although the autocorrelation still is very high and greater than for AE or ap, the solar wind bulk flow, as represented by B, begins to dominate the other correlators. These variations, when interpreted in the light of the basic solar wind structure, appear to be consistent with current ideas.

## REFERENCES

- Arnoldy, R. L. (1971): Signature in the interplanetary medium for substorms. J. Geophys. Res., 76:5189.
- Ballif, J. R., D. E. Jones, and P. J. Coleman, Jr., (1969): Further evidence of the correlation between transverse fluctuations in the interplanetary magnetic field and Kp. J. Geophys. Res., 74:2289.
- Ballif, J. R., D. E. Jones, P. J. Coleman, Jr., L. Davis, and E. J. Smith (1967): Transverse fluctuations in the interplanetary magnetic field: a requisite for geomagnetic variability. J. Geophys. Res., 72:4357.
- Bevington, P. R. (1969): Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill Book Company, New York.
- Bobrov, M. S. (1973): Kp index correlations with solar-wind parameters during the first and second stages of a recurrent geomagnetic storm. Planet. Space Sci., 21:2139.
- Burton, R. K., R. L. McPherron, and C. T. Russell (1975): The terrestrial magnetosphere: A half-wave rectifier of the interplanetary electric field. Science, 189:717.
- Crooker, N. V., J. Feynman, and J. T. Gosling (1977): On the high correlation between long-term averages of solar wind speed and geomagnetic activity. J. Geophys. Res., 82:1933.
- Dessler, A. J. (1967): Solar wind and interplanetary magnetic field. Rev. Geophys. Space Phys., 5:1.
- Diodato, L., G. Moreno, C. Signorini, and F. W. Ogilvie (1974): Long-term variations of the solar wind proton parameters. J. Geophys. Res., 79:5095.
- Draper, N. R., and H. Smith (1966): Applied Regression Analysis. John Wiley and Sons, Inc., New York.
- Foster, J. C., D. H. Fairfield, K. W. Ogilvie, and T. J. Rosenberg (1971): Relationship of interplanetary parameters and occurrence of magnetospheric substorms. J. Geophys. Res., 76:6971.
- Garrett, H. B. (1974): The role of fluctuations in the interplanetary magnetic field in determining the magnitude of substorm activity. Planet. Space Sci., 22:111.



- Garrett, H. B., D. F. Smart, and M. A. Shea (1978): A study of the use of interplanetary magnetic field and plasma measurements as a predictor of geomagnetic activity. In: Contributed Papers Presented at the Solar-Terrestrial Physics Symposium. Innsbruck, 1978 (Magnetospheric): 165.
- Garrett, H. B., T. W. Hill, and A. J. Dessler (1974): Influence of solar wind variability in geomagnetic activity. J. Geophys. Res., 79:4603.
- Hirshberg, J., and D. S. Colburn (1969): Interplanetary field and geomagnetic variations - a unified view. Planet. Space Sci., 17:1183.
- Iyemori, T., and H. Maeda (1979): Prediction of geomagnetic activities from solar wind parameters based on linear prediction theory. In: Proceedings of the International Solar-Terrestrial Predictions Workshop, Paper 18.
- King, J. H. (1975): Interplanetary Magnetic Field Data Book. NASA/NSSDC 75-04. NASA, Greenbelt, Maryland.
- King, J. H. (1977): Interplanetary Medium Data Book. NASA/NSSDC 77-04. NASA, Greenbelt, Maryland.
- Lenhart, K. G. (1968): Geomagnetic and solar data for use with digital computers. Trans. A.G.U., 49:463.
- Maezawa, K. (1978): Dependence of geomagnetic activity in solar wind parameters: A statistical approach. Solar Terrestrial Environmental Research in Japan, 2:103.
- Perreault, P., and S.-I. Akasofu (1978): A study of geomagnetic storms. Geophys. J. Roy. Astron. Soc., 54:547.
- Rostoker, G. (1972): Geomagnetic indices. Rev. Geophys. Space Phys. 10:935.
- Schatten, K. H., and J. M. Wilcox (1967): Response of the geomagnetic activity index Kp to the interplanetary magnetic field. J. Geophys. Res., 72: 5185.

## B. PREDICTION OF ENERGETIC PARTICLE DISTURBANCES

A WORKING GROUP REPORT by G. A. Paulikas, Chairman, D. N. Baker, W. R. Barron, V. Domingo, P. R. Higbie, W. L. Imhof, L. R. Lyons, R. L. McPherron, E. C. Roelof, M. Scholer, M. A. Shea, D. F. Smart, W. N. Spjeldvik and J. I. Vette.

### 1. INTRODUCTION

The Working Group on Energetic Particle Disturbance Prediction evaluated the present status of our capabilities to predict energetic particle disturbances in space and developed a series of findings and recommendations. These recommendations, if implemented, promise to significantly increase the accuracy of existing forecasting techniques for the prediction of solar particle fluxes near the earth and to allow the development of predictive techniques for forecasting magnetospheric particle disturbances - a field which, at the present time, totally lacks operational short-term forecasting capability.

This report is structured to emphasize our findings and recommendations. The interested reader will find additional relevant background material in the Appendices attached to the Working Group report. We have also included a selected list of references, many of which can be found in these Proceedings. These references may illuminate various aspects of the state of the art of energetic particle disturbance predictions or serve as authoritative reviews of work in this field.

### 2. SCOPE

This report specifically addresses the prediction of fluxes of energetic particles in space. Necessarily, our interests overlap those of other Working Groups concerned with the predictions of solar, interplanetary and magnetospheric phenomena and the reader should consider our findings and recommendations in the context of the entire field of magnetospheric physics and solar flare predictions. We have chosen to limit the scope of our recommendations by, first, considering only how to predict energetic particle fluxes, be they of solar or magnetospheric origin. Second, we have limited our attention to those areas which are of major importance in a practical sense, i.e., where either operational prediction schemes exist (Avdyushin et al., 1979; Heckman, 1979) or seem to be highly desirable and where an improved predictive capability can be of significant benefit to solar-terrestrial physics as well as to applications-oriented users in society. Using these criteria as a filter, we chose to address in detail the problems of solar energetic particle prediction, the prediction of the properties of the particle populations generated by magnetospheric storms and substorms, and the prediction of long-term variations in the populations of magnetospheric particles. Detailed findings and recommendations for each of the areas are given below.

The sections below summarize our findings and recommendations. The recommendations concerning prediction of solar particles were developed with a view of building on the base of the operational PPS-76 program described elsewhere in these Proceedings by Smart and Shea (1979). The recommendations concerning predictions of long term particle variations are a natural outgrowth of the work on magnetospheric particle population modeling reviewed by Vette et al. (1979) in these Proceedings. These references and the reports of

the other Working Groups on solar and magnetospheric predictions are a necessary (reading) precursor for the present report. We also recommend careful reading of the report prepared by the Working Group on spacecraft applications for the purpose of assessing such gaps as may presently exist between requirements for forecasts and the potential capabilities which exist and can provide quantitative forecasts of particle disturbances.

### 3. FINDINGS AND RECOMMENDATIONS

#### 3.1 Solar Flare Particles

##### 3.1.1. Warnings and Reaction-Mode Predictions

The short term warning mode - forecasting of the particle fluxes which may arrive at the earth from a flare before the flare occurs - requires the prediction of the gross optical, radio and X-ray emission characteristics of the flare. A lumped parameter, such as the "comprehensive flare index" (CFI) of Dodson and Hedeman (1971) might be used for a rather rough estimate of peak flux as well as the total  $> 10$  MeV proton flux and relativistic electron fluxes in cases where coronal and interplanetary propagation are expected to be about average. If predictions of the probability of flare occurrences (e.g., in terms of the distribution of the CFI) can be made for the entire transit of an active region, or for several solar rotations then rough estimates of energetic particle fluxes could be attempted in the same way for a long-term warning mode. For predictions of particle fluences or other gross parameters describing solar particles over a complete solar cycle the best course would be to use retrospective data compilations covering the last two solar cycles (Modisette et al., 1965; King, 1974; Stassinopoulos and King, 1972) to estimate the likely course of particle emissions during future solar cycles.

The short-term reaction mode - forecasting the subsequent history of a particle event after the flare has occurred - reduces many of the ambiguities of the warning mode by estimating from real-time data:

- (i) Particle energy spectrum and composition at the acceleration site
- (ii) Spatial and temporal dependence of particle emission from the corona
- (iii) Propagation from the corona to earth

We analyze these three elements of the short-term reaction mode in detail below, and then point out how some of the techniques can be applied to improving the short-term warning mode. The reader should note that a program called PPS-76, developed by Smart and Shea and described elsewhere in these Proceedings is in operational use and already incorporates some of the ideas which are suggested as requiring further development below. The present discussion builds on the experience and insight gained from developing and using PPS-76.



### 3.1.2. Spectrum and Composition

Currently the most reliable predictor of  $> 10$  MeV proton fluxes (e.g., PCA events) is the flare radio burst spectral distribution at millimeter and centimeter wavelengths (Castelli and Guidice, 1976; Akin'yan et al., 1977a, 1979a, 1979b; Bezrchenkova et al., 1977). The predictions of proton intensities have been improved by incorporating rough parameters describing coronal propagation. The slope of the proton spectra can also be predicted with reasonable accuracy from radio observations (Bakshi and Barron, 1979a, b; Barron and Bakshi, 1979). There have been few attempts to predict the intensities and spectra of relativistic electrons since the relativistic electrons usually accompany protons  $> 10$  MeV and their spectra are rather reproducible ( $\frac{dN}{dE} \propto E^{-3}$ ). The prediction of relativistic ground level events (GLE) from flare diagnostics is less reliable than for subrelativistic energies.

Composition at high energies (e.g., the proton/helium ratio) is not too variable in large events, but the recently discovered low-energy ( $\sim 1$  MeV/nucleon) "Z-rich" events (Hovestadt et al., 1975; Anglin et al., 1977; Zwickl et al., 1978) are usually from small flares on the sun's western hemisphere and are consequently difficult to predict.

The use of satellite measurements of soft X-ray fluxes ( $\sim 1-10$  A) for prediction have not been exploited to the extent that the radio observations have. However, the significant contribution of X-ray related indicators to the mix of parameters in the comprehensive flare index (CFI) of Dodson and Hedeman (1971) suggests that the soft X-ray signature may be a good candidate as an additional parameter which may improve our abilities to characterize the acceleration processes occurring in flares.

#### Recommendations

- (i) Multi-frequency radio patrols (24 hour coverage) should be continued and the data fed to forecasters for the purpose of improving reaction mode prediction of flare proton intensities and spectra.
- (ii) Satellite measurements of soft ( $\sim 1-10$  A) solar X-rays should continue, and there should be further scientific investigation of the use of both soft and hard X-rays as proton and electron prediction parameters.
- (iii) Optical flare patrols (24 hour coverage) should be continued as these still provide the basic alert for possible flare particle events.

### 3.1.3 Coronal Particle Emission

#### Findings

Although we do not know whether particle accumulation into coronal storage is impulsive or extended in duration, we do know from satellite and ground-based measurements of particle flux anisotropies that  $> 10$  MeV protons and  $> 0.2$  MeV electrons have been continuously released for as long as a

day after a large flare (Roelof and Krimigis, 1977). There is no known solar diagnostic of extended injection; the information regarding extended injection is inferred indirectly from the large outward field-aligned anisotropies of the particles observed in interplanetary space.

Once injected into the corona, high energy particles can cover virtually all longitudes in the corona, and the emission of protons and ions from the corona can be extremely heterogeneous. In the two largest flare proton events in the last solar cycle, multiple spacecraft observations using near-earth and Pioneer detectors revealed intensity differences in  $> 10$  MeV protons of more than a factor of 100 across  $> 50^\circ$  in longitude a day or more after the flare (Keath, et al., 1971; Roelof et al., 1974). Coronal magnetic structure must hold the key to this behavior. A coarse diagnostic of global coronal structure is the H $\alpha$  Synoptic Chart (McIntosh, 1972; 1979) which delineates large-scale magnetic polarity boundaries, and has been applied with some success to particle event prediction (Gold and Roelof, 1976; Roelof et al., 1977; Akin'yan and Chertok, 1977; Akin'yan et al., 1977b). More detailed associations of coronal magnetic structures with particle transport have been made using potential-field calculations based on high resolution Kitt Peak magnetograms, but the most direct identifications are possible from emission loops visible in the soft X-ray (44-54 A) and EUV images obtained from Skylab and various rocket shots (Roelof, 1979).

Although not yet fully scientifically validated, it appears that the non-homogeneous transport could result from the requirements that there be closed loops to move the accelerated particles across the corona, but that these particles must eventually find regions of predominantly open field lines in order to escape into the interplanetary medium. These open regions then would be the structures that modulate particle injection into interplanetary space as a function of solar longitude and latitude.

#### Recommendations

- (i) Acquire one image per day of the sun in soft X-rays (  $\sim 40\text{\AA}$ ) or EUV with  $\approx 5''$  resolution (sufficient to identify emission loops  $\approx 0.2R_\odot$  above the photosphere). White light coronagraph images are considerably less useful because they do not directly define structures on the disk without extensive and ambiguous deconvolution.
- (ii) Continue of daily coverage of the sun with high resolution full disk magnetograms.
- (iii) Develop daily "update" calculations of potential-field magnetic structure over limited or entire photosphere (the latter requiring the previous 27 days of observations for one global computation).
- (iv) Continue scientific analysis of coronal particle transport. Paragraphs (i-iii) of this recommendation can be implemented to provide operationally useful data to forecasters.



### 3.1.4. Interplanetary Propagation

#### Findings

During the rising phase of most flare particle events, the flux of particles is anisotropic and field-aligned, indicating that particles faithfully follow (moving) interplanetary field lines (Roelof, 1979); this lack of cross-field scattering should also hold in the decay phase of the event (after coronal injection is essentially over), when the particles are being "convected" out of the inner heliosphere by the moving field lines (Zwickl and Roelof, 1979). Once the event maximum has been properly identified, the decay phase can be modeled rather accurately, as long as the field lines at earth have sampled a relatively homogeneous coronal injection history. This is often not the case because of strong injection gradients in the corona, so there may be abrupt rises or drops during the decay phase. There are also abrupt changes during the rise phase because our interplanetary field connection to the corona traverses the transition between regions with vastly different emission rates. Changes in the coronal connection cause the principal distortion in solar event histories.

Since the solar wind velocity structure controls the evolution of the large scale interplanetary field, it is fortunate that knowledge of the instantaneous solar wind velocity allows us to estimate the coronal connection point of the field line through the spacecraft at that time (Nolte and Roelof, 1973a, b). Real-time solar wind measurements therefore allow prediction (in the reaction mode) of abrupt discontinuities in particle fluxes. On the other hand, estimates of evolving solar wind structure would allow a warning mode prediction of abrupt flux changes. Such estimates could come from images of coronal structure (if the modulation is caused by a co-rotating stream) or estimates of the arrival of flare-generated plasma disturbances (which require theoretical calculations that realistically model the three-dimensional distortions of the interplanetary field).

The most promising, immediately available "remote sensing" technique for the solar wind is the interplanetary scintillation (IPS) of galactic and extragalactic radio sources of small angular diameter (Watanabe, 1979). The main limitation of currently operating IPS arrays is the relatively small number of reliably scintillating sources ( $< 10$ ) at observing frequencies  $> 70$  MHz. Only a few of these sources lie in the ecliptic plane, so directional sensitivity to approaching solar wind disturbances has a seasonal dependence. Nonetheless, IPS measurements during some months provide unique and valuable information for warning mode predictions of strong distortions in particle event histories.

#### Recommendations

- (i) Solar wind speed should be provided in real time for prediction of particle event histories in the reaction mode. The paper of Tsurutani and Baker (1979) in these Proceedings describes steps presently being taken to implement this recommendation.
- (ii) Interplanetary scintillation multi-site observations of



solar wind speed should be obtained on a daily basis. Even though IPS directional sensitivity is seasonal, measurements during the optimal periods of the year are extremely valuable for predictions in the warning mode.

- (iii) Research concerning three-dimensional distortion of the interplanetary field by solar flare plasma disturbances is essential to the understanding necessary for predicting the effect of these disturbances on energetic particle event histories.

### 3.2 Magnetic Storms

A magnetic storm is an interval of several hours to several days duration during which the horizontal component of the magnetic field at a near equatorial ground station undergoes a significant deviation from quiet values. A storm is characterized by the  $D_{st}$  index, which is an average over all local times of the deviations in  $H$ , and the asymmetry index, i.e., the range between maximum and minimum values of the deviation in  $H$ . These indices represent and are related to the currents of trapped particles circulating in the magnetosphere. The trapped particle populations in the entire outer magnetosphere undergo massive changes during and after magnetic storms with resulting ionospheric and aeronomic effects. The reader should consult the reports of the Working Groups on Interplanetary-Magnetosphere Interactions and Geomagnetic Disturbance Predictions to become more familiar with the phenomenology of magnetic storms and the report of the Working Groups dealing in ionospheric predictions and spacecraft applications to appreciate the effects.

There are two types of magnetic storms, those associated with solar flares and recurrent storms which are most probably associated with coronal holes. The geoefficiency of a flare depends on the location of the flare, its size and the direction of the photospheric magnetic field. Recurrent magnetic storms are more common near solar minimum. Their effectiveness in generating geomagnetic activity appears to be a consequence of an interaction between a high speed solar stream and slower plasma ahead of it (Mishin et al., 1979; Iyemori and Maeda, 1979). The interaction organizes the various plasma and field parameters in a systematic way. In particular, the interplanetary magnetic field (IMF) is tipped out of the ecliptic plane with a sudden transition from north to south, (or vice versa) occurring at the center of the interaction region.

Magnetic storms are controlled by the direction of the IMF and the solar wind velocity through the interplanetary electric field,  $E_{sw} = -VB_s$  where  $B_s$  is equal to the GSM Z component of the IMF, if it is negative, and is zero otherwise. When the magnetosphere is treated as a linear system it is possible to predict the  $D_{st}$  index as a function of time using an impulse response of roughly 10 hours  $D_{st}$  duration convolved with  $E_{sw}$ . Thus, in principle, given the parameters of the interplanetary medium it is possible to predict  $D_{st}$ .

### Recommendation

We recommend that a workshop be organized to test and evaluate our ability to predict  $D_{st}$ . Data describing the state of the interplanetary medium would be used to predict  $D_{st}$  and then compared to  $D_{st}$  values calculated from ground observations.

Several processes are thought to be responsible for the injection and energization of ring current particles. A sudden enhancement of an azimuthal electric field will cause ambient particles drifting on closed paths inside the plasmopause to move inward. As they do this they gain energy and drift more rapidly creating an enhanced ring current (Lyons and Williams, 1979). Simultaneously particles at greater distances in the magnetotail drift earthward gaining energy and replacing those previously present and accelerated out of the region. Sporadically, substorms divert a portion of the tail current down magnetic field lines westward through the electrojet. This causes rapid increases in the magnetic field in a localized sector near midnight. This in turn creates an inductive electric field which accelerates particles through the betatron and drift-betatron processes. These particles rapidly drift out of the acceleration region becoming part of the source which supplies the inner magnetosphere (Clauer and McPherron, 1979).

### Recommendation

We recommend additional coordinated research or workshop-type activities which will better describe the quantitative relationship between  $D_{st}$ , or other magnetospheric parameters and indices, and the actual energy spectra and spatial distributions of the storm time ring current particle population.

The decay of the ring current is via wave particle interactions and charge exchange. When the IMF turns northward, the plasmopause begins to expand. As it expands, both the proton and electron distribution functions become cyclotron unstable producing waves which scatter particles into the loss cone thus precipitating the particles into the ionosphere.

The rate of precipitation of energetic electrons can be calculated from knowledge of the spectrum of plasmaspheric whistler-mode hiss (Spjeldvik and Lyons, 1979). Particles precipitated by this process penetrate to the D region of the ionosphere enhancing the ionization of this layer. This ionization alters atmospheric chemistry producing such effects as changes in the ozone content as well as effects on communications which can be quantitatively predicted.

The various physical processes linking a solar flare or coronal hole to eventual particle precipitation into the atmosphere are thought to be understood to varying degrees. By combining a variety of semiquantitative empirical relations and theoretical calculations, it should now be possible to crudely predict the magnetospheric effects of storms and the resultant atmospheric effects of particle precipitation during magnetic storms. An essential element for such predictions is measurement of the properties of the solar wind plasma and magnetic field immediately upstream of the earth's bow shock.



### Recommendation

We recommend that a coordinated workshop be organized, involving magnetospheric and atmospheric scientists, for the purpose of assessing, using selected storm time data, the feasibility of predicting the beginning-to-end evolution of one or more magnetic storms, including aeronomic and ionospheric effects.

### 3.3 Substorms

At the present time it is difficult to agree on precise interpretations of all of the detailed features of magnetospheric substorms. However, associated with all significant substorm activity there are major changes in the near-earth magnetic field and there are also substantial re-distributions and energizations of particles in the outer magnetosphere. We conclude that major substorm effects involve a release of a large fraction of the energy which has been extracted from the solar wind and which has been stored within the magnetosphere. We adopt, therefore, the point of view that it is, and ought to be, of primary interest to users to define substorms in terms of the major substorm energy release in the form of charged particles. The reference substorm onset time ( $t=0$ ) can be specified as the time of first appearance of intense fluxes of energetic magnetospheric particles near local midnight.

Warning Mode - Since we wish to predict substorm energetic particle disturbances, we presently recognize two powerful tools that are available for this purpose:

- 1) Measurement of the solar wind energy input function,  $f(V_{sw}, B_z)$ , to the magnetosphere.

This basically involves some combination of the solar wind speed ( $V_{sw}$ ) and the IMF north-south component ( $B_z$ ). A measurement of  $f(V_{sw}, B_z)$  can tell us the size of magnetospheric effects (substorms), but not when or where effects will occur. It seems possible that lead times of 30-60 minutes may be achieved if measurements were made at the sunward libration point (Tsurutani and Baker, 1979).

- 2) Measurements of the internal stresses of the magnetosphere.

Such measurements can be made using energetic electron anisotropies and/or sensitive magnetometers at geostationary orbit (Baker et al., 1979a, b). This technique gives a fairly good indication (on a statistical basis) of when a substorm will occur. Substorm warnings one to two hours in advance of substorm onset times may be provided by the method.

The size of the open field line region over the polar cap as determined by low altitude spacecraft or ground based magnetometers or radars is also a measure of the stress in the magnetosphere.



### Recommendations

- (i) Evaluate, in an operational sense, the efficacy of using the actual magnetospheric energy input function,  $f(V_{sw}, B_z)$  as measured by ISEE-3 interplanetary data to compute the energy dumped by substorms using a suitable index.
- (ii) Guided by result of the recommendation above, implement a real-time ISEE-3 data system. The paper by Tsurutani and Baker in these Proceedings provides a roadmap for acquiring the ISEE-3 data.
- (iii) Develop a real-time system for utilization of geostationary energetic electron data for predicting substorm onsets and utilize the magnetometers presently in the geostationary orbit to aid in the measurements of magnetospheric stresses.
- (iv) Using the appropriate data sets, study the details of magnetic field direction changes and electron drift-shell changes which might be used to estimate the size of the impending substorm disturbance.

Reaction Mode - Once a substorm has taken place, what can be done to update predictions or to specify the magnetospheric effects?

- 1) Magnetospheric effects (important for spacecraft charging or spacecraft operations):
  - a) Plasma parameters can be measured directly by plasma analyzers or inferred by appropriate fitting techniques using data from more limited instrumentation.
  - b) Radiation dose measurements can be made which are useful for assessing hazards to spacecraft operations.
- 2) Ionospheric and auroral effects (important for HF communications):
  - a) Location and extent of auroral zone can be inferred from low altitude measurements of electron precipitation, or bremsstrahlung X-rays, or high altitude optical and X-ray imaging.
  - b) Sensing of auroral zone precipitation from synchronous orbit observations may be possible for some applications through measurement of
    - i) Strong pitch angle scattering
    - ii) Prediction of drift trajectories in the magnetosphere using realistic models of the magnetosphere.

### Recommendation

We recommend that research be undertaken to develop global models of the magnetospheric magnetic field fitted to real time measurements of magnetic field components (or the loss cone direction of energetic particle distributions) made by a number of satellites. Encouragement should be given to develop instruments capable of imaging the auroral zone in UV or X-rays. Further effort should be devoted to study of parallel electric fields which considerably alter particle distribution functions at low altitudes and strongly affect auroral features; such efforts are clearly essential if low altitude effects are to be inferred from high altitude measurements.

### Phenomena Not Predicted:

There are, obviously, many phenomena of interest to the scientific and application community, which we cannot predict. Several examples will suffice.

- 1) Ionospheric effects whose relationship to substorms is not clear or obvious. These include relativistic electron precipitation (REP) events and precipitation spikes observed by low altitude spacecraft.
- 2)  $\frac{\partial B}{\partial t}$  during substorms (of interest for their inductive effects) except that magnitude of  $B$  is probably related to size of energy input  $f(V_{sw}, B_z)$ . Detailed knowledge of the substorm process and its triggering mechanism must be known before such estimates can be made.

### Recommendation

Users should correlate times of observed effects in their systems with commonly used definitions of substorm onsets (particle injection times, mid-latitude positive bays in magnetograms) so that forecasters and scientists can know more precisely what is important to users. Publication of monitoring data might be of service in this regard.

## 3.4 Long Term Variations of Energetic Particles

In a predictive or warning mode, the long term averaged quantitative models of trapped protons and electrons described by Vette et al. (1979) in these Proceedings can be used to provide the fluences received by any spacecraft orbiting within the limits of the model. There is one region of space between  $L$  of 2 and 7 where predictions are not reproducible from year to year within acceptable accuracy. Solar proton fluence models consisting of event integrated fluences and statistics over the 20th solar cycle have been prepared and provide a good first-order estimate of the likely solar proton fluxes expected during the 21st cycle.



### Recommendations

- (i) Maintain quantitative trapped radiation electron and proton models to an accuracy of a factor of about 2 by incorporating new data at appropriate times.
- (ii) Following the completion of the 21st solar cycle, compile the event integrated fluences and statistics for the solar proton events from 5 MeV up to the maximum energies observed and using riometer absorption events, recalibrate the 19th solar cycle compilations so that three cycles will be available to predict fluences in the 22nd cycle and beyond. The efforts to maintain a high accuracy historical record of galactic cosmic ray flux should continue.
- (iii) Determine a representation of the electron fluxes that appear in the trapped region following a magnetic storm characterized by the  $D_{st}$  parameter and give its decay features over the whole region; this should be done in the frame work of a multi-storm sequence. In a reaction or specification mode there are several products that can be produced with existing data and the cause of the variability in multi-month flux averages over a portion of the outer zone can likely be determined.
- (iv) On the basis of data collected over the 20th solar cycle, compile the statistics on storm injected electrons appropriate to planning a work schedule for man operating at the geostationary orbit for large facility construction and maintenance or repair of outer zone spacecraft by Shuttle visitation.
- (v) Determine the cause of the time variations in outer zone electron fluxes characterized by other than the exponential decay, particularly those decays associated with flux increases produced by fast solar wind streams impacting the magnetosphere.
- (vi) In order to speed up the pace of quantitative model development and to permit the experimenters to participate directly in this activity, use the newly developed coordinated data analysis workshop technique, in which computer interactive graphics is employed to manipulate and display a common data base constructed to address a specific problem.



4. SUMMARY OF RECOMMENDATIONS:  
WORKING GROUP ON ENERGETIC PARTICLE DISTURBANCES

<u>Recommendation</u>	<u>Prediction</u>	<u>Status of Pre- diction Technique</u>	<u>Implications</u>
<u>Solar Particles</u>			
<u>Solar active region evolution</u> (See Report of Working Groups on Solar Activity Prediction)			
<u>Prediction of Particle Spectrum and Composition</u>			
(i) Multi-frequency radio patrols	Proton intensity and spectrum	Operational	Operation of several ground stations
(ii) Soft (1-10A) solar x-ray measurement	Proton and electron prediction	Research	Satellite in orbit
(iii) Optical flare patrol	Particle flare occurrence	Operational	Operation of several observatories
<u>Prediction of Coronal Transport</u>			
(i) Soft x-rays (40 A) or EUV solar imaging 1/day	Coronal loop structure	Capability exists	Satellite in orbit
(ii) High resolution solar disk magnetograms	Magnetic field sector structure	Operational	Ground based observations
(iii) Daily calculation of potential field B structure over photosphere	Coronal loop structure	Research	Research
(iv) Research of coronal particle transport	2 Dimen- sional profile of particle flux release from the solar corona	Research	Research
<u>Prediction of Interplanetary Transport</u>			
(i) Solar wind speed in real time	Particle event devel- opment	Capability exists	Satellite in orbit
(ii) Multi-site interplanetary scintillations	Particle event devel- opment/prediction	Operational	Increase ground observations

(iii) 3-dimensional distortion of interplanetary magnetic field by solar activity	Effect of solar activity on particle flux profiles	Research	Research
---	--	----------	----------

#### Magnetic Storms

(See Reports of Working Groups on Interplanetary-Magnetosphere Interactions and Geomagnetic Disturbance Predictions)

Workshop to evaluate ability to predict $D_{st}$ using interplanetary medium data	Prediction of magnetic storms	Research	Coordinated research
---	-------------------------------	----------	----------------------

Research and relationship between $D_{st}$ or other magnetospheric indices and the storm time ring currents	Prediction of particles in magnetosphere during magnetic storms	Research	Coordinated research
---	---	----------	----------------------

Coordinated Workshop (involves magnetospheric and atmospheric scientist) to study the feasibility of predicting end-to-end evolution of one or more magnetic storms	Prediction of magnetic storm effects	Proposed	Coordinated research
---	--------------------------------------	----------	----------------------

#### Magnetospheric Substorms

Evaluation of magnetospheric input function with respect to the Solar wind-data provided by ISEE-3	Capability to predict substorm activity from solar wind data	Research	Research
--	--	----------	----------

Implement real-time ISEE-3 data system	Substorm activity/warning 30-60 min	Capability exists	Data handling system
--	-------------------------------------	-------------------	----------------------

Develop real-time system for geostationary electron data prediction of substorm onset	Substorm activity/warning 30 min	Capability exists	Data handling system
---	----------------------------------	-------------------	----------------------

Research on predictive characteristics of B direction changes and electron drift-shell changes	Substorm activity/warning 1-2 hours	Research	Research
--	-------------------------------------	----------	----------

Research to develop models of the magnetospheric B, Auroral zone UV and X-ray imaging, parallel electric field	Low altitude effects	Research	Instrumentation for auroral zone X-ray and UV Imaging
--	----------------------	----------	---

User effects correlations to be done with substorm onsets	User effects prediction development	Proposed	Research
<u>Long term variations of energetic particles</u>			
Trapped radiation e and p models to be maintained	Particle fluxes in trapped radiation region	Operational	Data handling
Compilation of integrated fluxes of solar events ( $E_p > 5$ MeV) and of galactic cosmic rays	Particle fluxes over polar caps and outside the magnetosphere	Operational	Data handling
Model of electron flux in trapped region following magnetic storm in function of $D_{st}$	Electron fluxes in outer radiation zone	Proposed	Research and data handling
Storm injected electrons in outer radiation zone: statistical compilation	Electron fluxes in geostationary orbit	Proposed	Research and data handling
Research on time variations of outer zone electron fluxes, associated with solar wind	Electron flux in outer zone	Research	Research
Coordinated data analysis workshops		Proposed	Coordinated research

## 5. REFERENCES

- Akin'yan, S. T., I. M. Chertok, (1977): Determination of the Parameters of Solar Protons in the Vicinity of Earth from Radio Bursts, 3. Temporal reference functions. Geomagnetism and Aeronomy 17, 407.
- Akin'yan, S. T., V. V. Fomichev and I. M. Chertok (1977a): Determination of the Parameters of Solar Protons in the Neighborhood of the Earth from Radio Bursts, 1. Intensity functions. Geomagnetism and Aeronomy 17, 5.
- Akin'yan, S. T., V. V. Fomichev and I. M. Chertok (1977b): Determination of the Parameters of Solar Protons in the Vicinity of the Earth from Radio Bursts, 2. Longitudinal attenuation functions. Geomagnetism and Aeronomy 17, 123.



- Akin'yan, S. T., I. M. Chertok and V. V. Fomichev (1979a): Quantitative Diagnostics of Solar Proton Flares by Radio Data. These Proceedings.
- Akin'yan, S. T., I. M. Chertok and E. M. Zhulina (1979b): Determination of PCA Value by Characteristic of Solar Radio Bursts. These Proceedings.
- Anglin, J. D., W. F. Dietrich and J. A. Simpson (1977): Super Enrichments<sub>3</sub> of Fe Group Nuclei in Solar Flares and their Association with Large <sup>3</sup>He Enrichments. Proceedings 15th International Cosmic Ray Conference (Munich) 5, 43.
- Avdyushin, S. I., N. K. Pereyaslova, F. L. Deikman and Yu. M. Kulagin (1979): Forecasting of Radiation Situation in the Forecast Center of the Institute of Applied Geophysics. These Proceedings.
- Baker, D. N., R. D. Belian, P. R. Higbie and E. W. Hones, Jr. (1979a): Prediction of High Energy ( > 0.3 MeV) Substorm Related Magnetospheric Particles. These Proceedings.
- Baker, D. N., P. R. Higbie, E. W. Hones, Jr., and R. D. Belian (1979b): The Use of > 30 KeV Electron Anisotropies at 6.6 R<sub>e</sub> to Predict Magnetospheric Substorms. These Proceedings.
- Bakshi, P. and W. Barron (1979a): Prediction of Solar Flare Proton Spectral Slope from Radio Burst Data. J. Geophys. Res. 84, 131.
- Bakshi, P. and W. R. Barron (1979b): Predictions of Solar Flare Proton Spectrum from Radio Burst Characteristics. These Proceedings.
- Barron, W. R. and P. Bakshi (1979): Application of Integrated Radio Burst Fluxes to the Prediction of Solar Energetic Proton Flux Increases. These Proceedings.
- Bezrhenkova, T. M., N. A. Mikryukova, N. K. Pereyaslova and S. G. Frolov (1977): Identification of Solar-Flare Protons from the Accompanying Electromagnetic Emission. Geomagnetism and Aeronomy 17, 544.
- Castelli, J. P. and D. A. Guidice (1976): Impact of Current Solar Radio Patrol Observations. Vistas in Astronomy 19, 355.
- Cauffman, D. P. (1979): Space Plasma Monitoring: Two New Capabilities for the Coming Decade. These Proceedings.
- Clauer, C. R. and R. L. McPherron (1979): Predicting Partial Ring Current Development. These Proceedings.
- Dodson, H. W., E. R. Hedeman and O. C. Mohler (1979): Examples of "Problem" Flares or Situations in Past Solar-Terrestrial Observations. These Proceedings.
- Dodson, H. W. and R. Hedeman (1971): An Experimental, Comprehensive Flare Index and its Derivation for "Major" Flares, 1955-1969. Rep. UAG-14, World Data Center - A, NOAA (Boulder, Colo.).

- Gold, R. E. and E. C. Roelof (1976): A Prediction Technique for Low Energy Solar Proton Fluxes Near 1 AU. Space Research XVI, ed. M. J. Rycroft and R. D. Reasenberg, Akademie-Verlag (Berlin), 791.
- Heckman, G. R. (1979): A Summary of the Indices and Predictions of the Space Environment Services Center. These Proceedings.
- Hovestadt, D., B. Klecker, O. Vollmer, G. Gloeckler and C. Y. Fan (1975): Heavy Particle Emission of Unusual Composition from the Sun. Proc. 14th Int. Cosmic Ray Conf. (Munich) 5, 1613.
- Iyemori, T. and H. Maeda (1979): Prediction of Geomagnetic Activities from Solar Wind Parameters Based on the Linear Prediction Theory. These Proceedings.
- Keath, E. P., R. P. Bukata, K. G. McCracken and V. R. Rao (1971): The Anomalous Distribution in Heliocentric Longitude of Solar Injected Cosmic Radiation. Solar Phys. 18, 503.
- King, J. H. (1974): Solar Proton Fluences for 1977-83. J. Spacecraft and Rockets 11, 401.
- Lyons, L. R. and D. M. Williams (1979): A Source for the Geomagnetic Storm Main Phase Ring Current. To be published, J. Geophys. Res.
- McIntosh, P. S. (1972): Solar Magnetic Fields Derived from Hydrogen Alpha Filtergrams. Rev. Geophys. Space Phys. 10, 837.
- McIntosh, P. S. (1979): Annotated Atlas of  $H_{\alpha}$  Synoptic Charts for Solar Cycle 20, 1964-1974. World Data Center A, Special Report UAG-70, (NOAA/EDIS).
- Mishin, V. M., V. V. Shelomentsev, A. D. Bazarzhapov and L. P. Sergeeva (1979): On the Possibility for Short-Term Forecasting of Geomagnetic Storms Associated with High Speed Solar Wind Streams. These Proceedings.
- Modisette, J. L., T. N. Vinson and A. C. Hardy (1965): Model Solar Proton Environment for Manned Spacecraft Design. NASA Document TN D-2746.
- Nolte, J. T. and E. C. Roelof (1973a): Large-Scale Structure of the Interplanetary Medium. I: High Coronal Source Longitude of the Quiet-Time Solar Wind. Solar Phys 33, 241.
- Nolte, J. T. and E. C. Roelof (1973b): Large-Scale Structure of the Interplanetary Medium. II: Evolving Magnetic Configurations Deducted from Multi-Spacecraft Observations. Solar Phys. 33, 483.
- Paulikas, G. A. (1974): Tracing of High-Latitude Magnetic Field Lines by Solar Particles. Rev. Geophys. and Space Physics 12, 117.
- Roelof, E. C. (1979): Solar Energetic Particles: From the Corona to the Magnetotail. In "Quantitative Modeling of Magnetospheric Processes", ed. W. P. Olson, Geophysical Monograph 21, American Geophysical Union (Washington, D.C.), 220.



- Roelof, E. C. and R. E. Gold (1978): Prediction of Solar Energetic Particle Event Histories Using Real-Time Particle and Solar Wind Measurements. Operational Modelling of the Aerospace Propagation Environment, AGARD Conf. Proc. No. 238 1, VI-29, 1978.
- Roelof, E. C. and S. M. Krimigis (1977): Solar Energetic Particles Below 10 MeV, Study of Travelling Interplanetary Phenomena, 1977, ed. M. A. Shea, D. F. Smart and S. T. Wu, D. Reidel, (Dordrecht), 343.
- Roelof, E. C., J. A. Lezniak, W. R. Webber, F. B. McDonald, B. J. Teegarden and J. H. Trainor (1974): Relation of Coronal Magnetic Structure to the Interplanetary Proton Events of August 2-9, 1972. Correlated Interplanetary and Magnetospheric Observations (ed.) D. E. Page, D. Reidel (Dordrecht), 563.
- Roelof, E. C., R. E. Gold and E. P. Keath (1977): Evaluation of a Prediction Technique for Low Energy Solar Particle Events. Space Research XVII, ed. M. J. Rycroft and R. D. Reasenberg, Akademie-Verlag (Berlin), 545.
- Scholer, M. (1979): Energetic Solar Particle Behavior in the Magnetosphere. These Proceedings.
- Smart, D. F. and M. A. Shea (1979): PPS76-A Computerized "Event Mode" Solar Proton Forecasting Technique. These Proceedings.
- Spjeldvik, W. N. and L. R. Lyons (1979): On the Predictability of Radiation Belt Electron Precipitation into the Earth's Atmosphere following Magnetic Storms. These Proceedings.
- Stassinopoulos, E. G. and J. H. King (1972): An Empirical Model of Energetic Solar Proton Fluxes with Application to Earth-Orbiting Spacecraft. NASA/GSFC Document X-601-72-487.
- Tsurutani, B. T. and D. N. Baker (1979): A Real-Time ISEE Data System. These Proceedings.
- Vette, J. I., M. J. Teague, D. M. Sawyer and K. W. Chan (1979): Modeling the Earth's Radiation Belts. These Proceedings
- Watanabe, T. (1979): Solar-Terrestrial Predictions Using IPS Techniques. These Proceedings.
- Zwickl, R. D. and E. C. Roelof (submitted 1979): Interplanetary Propagation of  $< 1$  MeV Protons in Non-Impulsive Energetic Particle Events. J. Geophys Res. 84.
- Zwickl, R. D., E. C. Roelof, R. E. Gold, S. M. Krimigis and T. P. Armstrong (1978): Z-Rich Solar Particle Event Characteristics 1972-1976. Astrophys. J. 225, 281.



6. PARTICIPANTS IN THE WORKING GROUP ON ENERGETIC  
PARTICLE DISTURBANCE PREDICTIONS

D. N. Baker  
Los Alamos Scientific Laboratory  
Group P-4/MS 436  
Los Alamos, NM 87545

W. R. Barron  
AFGL/PHP  
Hanscom AFB  
Bedford, MA 01731

J. R. Burrows  
National Research Council  
Herzberg Institute of Astrophysics  
100 Sussex, R 2021  
Ottawa, CANADA

V. Domingo  
Space Science Department  
ESTEC  
Noordwijk, NETHERLANDS

Masashi Hayakawa  
The Research Institute of  
Atmospherics  
Nagoya University  
Toyokawa, 442 JAPAN

P. R. Higbie  
Los Alamos Scientific Laboratory  
Group P-4/MS 436  
Los Alamos, NM 87545

W. L. Imhof  
Dept. 52-12, Bldg. 205  
Lockheed Palo Alto Research  
Laboratories  
3251 Hanover Street  
Palo Alto, CA 94304

M. M. Kobrin  
Radiophysical Research Institute  
Lyadov Street 25/14  
603600 Gorkij, U.S.S.R.

L. R. Lyons  
NOAA/Space Environment Laboratory  
Boulder, CO 80302

R. L. McPherron  
University of California, Los Angeles  
Institute of Geophysics  
6847 Slichter Hall  
Los Angeles, CA 90024

L. I. Miroshinchenko  
IZMIRAN  
P/O Akademgorodok  
Troitsk  
Moscow Region 142092, U.S.S.R.

W. P. Olson  
MDAC/MS 13/3  
5301 Bolsa Avenue  
Huntington Beach, CA 92647

G. A. Paulikas, Chairman  
Space Sciences Laboratory  
The Aerospace Corporation  
P. O. Box 92957  
Los Angeles, CA 90009

E. C. Roelof  
Johns Hopkins University/Applied  
Physics Laboratory  
Johns Hopkins Road  
Laurel, MD 20810

M. Scholer  
Max-Planck-Institut Fur Physik  
Und Astrophysik  
Institut fur Extraterrestrische Physik  
8046 Garching b. Munchen  
GERMANY

M. A. (Peggy) Shea  
Space Physics Division  
AFGL/PHG  
Hanscom Air Force Base  
Bedford, MA 01731

D. F. Smart  
Space Physics Division  
AFGL/PHG  
Hanscom Air Force Base  
Bedford, MA 01731

W. N. Spjeldvik  
NOAA/Space Environment Laboratory  
Boulder, CO 80302

E. G. Stassinopoulos  
Code 601  
Goddard Space Flight Center  
Greenbelt, MD 20771

R. L. Thompson, Lt. Col., USAF  
Chief, Space Environmental Support  
Branch  
Department of the Air Force  
HQ Air Force Global Weather  
Central (MAC)  
Offutt Air Force Base, NEB 68113

A. L. Vampola  
Space Sciences Laboratory  
The Aerospace Corporation  
P. O. Box 92957  
Los Angeles, CA 90009

J. I. Vette  
Code 601  
Goddard Space Flight Center  
Greenbelt, MD 20904

D. J. Williams  
NOAA/Space Environment Laboratory  
Boulder, CO 80302

## APPENDIX I

### Short Term Magnetospheric Particle Variations ( $1 \text{ min} < T < 1 \text{ day}$ )

P. R. Higbie, D. N. Baker, V. Domingo, W. L. Imhof, R. L. McPherron  
W. N. Spjeldvik, D. J. Williams, J. R. Burrows, and M. Hayakawa

#### Introduction

The interest in studying short term ( $1 \text{ min} < T < 1 \text{ day}$ ) energetic magnetospheric particle variations stems from several sources, both scientific and applied. The scientific problems associated with this time domain include the entire substorm process as well as the coupling between the magnetosphere and interplanetary space (e.g., solar particle entry). The numerous practical problems which occur on the above timescales include spacecraft charging and radiation damage effects caused by solar proton events, communications interference from particle precipitation in the polar regions and even radiation hazards to people flying on great circle routes over the Arctic during solar proton events. Magnetospheric coupling to the atmosphere through precipitation of energetic particles gives rise to short term density disturbances in the thermosphere and is one link which has been suggested as joining the earth's weather and climate to solar activity. The ability to predict the occurrence, duration, evolution, and intensity of energetic particle flux variations in the magnetosphere is thus valuable both scientifically and for its contribution to practical human ventures.

#### Energetic Particle Processes

Figure 1 is a schematic representation of the generation and propagation processes for energetic particles of concern in solar-terrestrial predictions. The upper branch of the figure refers to particles energized at the sun by impulsive events such as flares. These particles travel through interplanetary space and eventually enter the Earth's magnetosphere and reach the atmosphere. The lower branch lists the processes by which particles are energized (0.01 - 5 MeV) in the earth's magnetosphere and then precipitate into the atmosphere or are otherwise lost from the magnetosphere. The processes listed in the lower branch include the substorm process which is still far from being completely understood. The reader should refer to the paper by Scholer in these Proceedings for a review of the solar particle entry process and to the reference in the reports of the Working Groups on Interplanetary-Magnetosphere Interactions and Geomagnetic Disturbance Predictions for additional insight into magnetospheric processes. Characteristic times are associated with various processes. Diamonds in the figure represent points at which one might break into the chain of events to make observations for predictive purposes.

The times that have been associated with the intermediate phenomena in the upper branch of Figure 1 are approximate characteristic times for 10 - 30 MeV/nucleon ions. For faster ions and for the electrons of interest, some of the quoted times are correspondingly shorter. We see that for any significant medium- and long-term forecast of solar energetic particle variations ( $> 1 \text{ day}$ ), only observations of the sun and the interplanetary medium for the purposes of predicting future particle emission and transport may be useful.

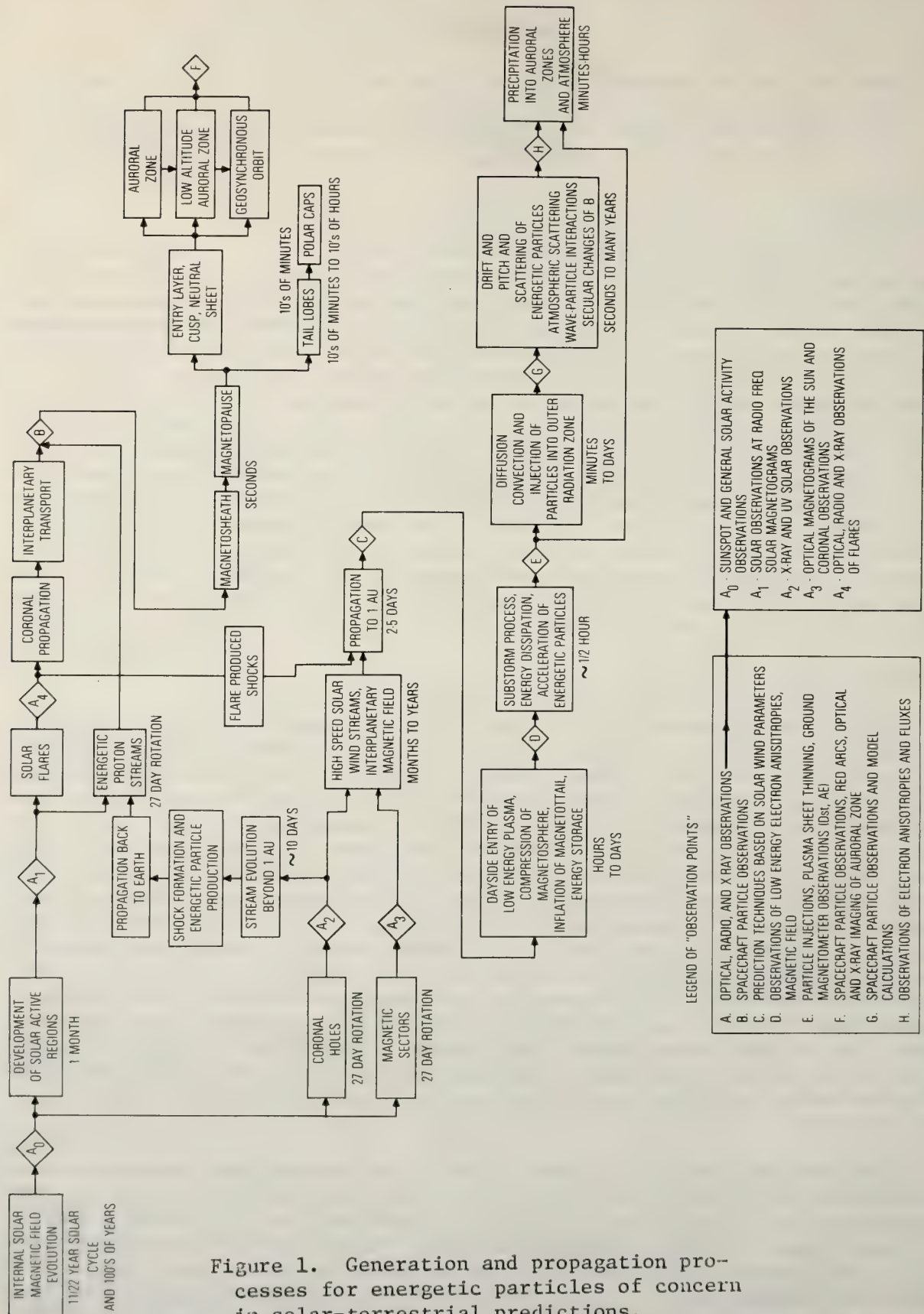


Figure 1. Generation and propagation processes for energetic particles of concern in solar-terrestrial predictions.



The reader should consult the reports of Working Groups on solar activity predictions to gain a perspective of the current status of our ability to predict solar activity both on the timescale of days as well as decades.

### Magnetosphere Penetration

Once particles of solar origin have reached the neighborhood of the magnetosphere, where they may be detected by satellites, the particles will enter the different magnetospheric regions with little delay.

A review of experimental results on solar particle penetration into the magnetosphere has been published by Paulikas (1974). The reader should also consult the paper by Scholer in these Proceedings for a more recent review. The major points of these reviews are:

- solar electrons have immediate (within seconds) access to the magnetotail and polar caps;
- protons of  $E > 1$  MeV have access to the low latitude boundary of the polar cap (that coincides more or less with the auroral region) in a very short time (in general, minutes) after reaching the magnetosheath. On the other hand the population of the tail lobes seems to suffer delays of the order of tens of minutes.
- Lower latitudes, within closed magnetic field lines, are penetrated when the geomagnetic activity is high. The same is true for geosynchronous orbit in the case of low energy protons. Higher energy protons ( $E_p > 20$  MeV) reach the geosynchronous orbit at all times.
- The precipitation of protons of 1 MeV up to at least 300 MeV near the centers of the polar caps is subject to delays that vary between some minutes and several hours. The amount of delay depends on the orientation of the interplanetary magnetic field, while the flux intensities reached over the polar caps are dependent on the anisotropy of the interplanetary flux.
- Observations made by HEOS-2 in the high latitude magnetosphere show that the behavior of particles observed in the polar cap is the same as for those observed in the deep tail. Particle fluxes in the regions near the magnetopause and the neutral sheet are seen to closely follow the interplanetary proton flux; fluxes at the center of the tail lobe, on the other hand, follow interplanetary levels only after some delay.

### Local Production of Energetic Particles

The ultimate energy source for locally energized particles in the Earth's magnetosphere is again the sun although particles originating in the earth's ionosphere may play a very significant role in magnetospheric dynamics. The Earth's ionosphere forms a complex lower boundary for the magnetosphere and there is strong evidence that part of the plasma in the magnetosphere is supplied by the ionosphere.

Coronal holes are the source of high speed solar wind streams. The persistence of coronal holes and magnetic field structure on the sun provide a recurrent pattern of high speed streams and magnetic sectors every twenty-seven days. Various structures in the solar wind including flare-produced shocks are propagated to the earth in two to five days. These structures evolve during the process of propagation and by the time the structures reach the distance of Jupiter's orbit, shock structures form due to the interaction of the high speed streams with more slowly moving solar wind particles. The closure of forward and reverse shocks from successive shock pairs results in a corotating cavity centered on the sun. The existence of such a cavity reinforces possible twenty-seven day periodicities. Thus, one crude prediction applicable to magnetospheric phenomena is the persistence of twenty-seven day patterns in energetic particle behavior.

The solar wind impinging on the Earth's dipole distorts the field and confines it to form the magnetosphere and the magnetopause boundary. Waves generated in this process act upon the incoming plasma to form a stand-off shock in the solar wind. These macroscopic structures are the average effect of microscopic processes. The solar wind parameters (density, temperature, bulk velocity, frozen-in interplanetary magnetic field) are functions of position and time. For example, standard deviations in the hourly averaged IMF magnitude can be comparable to the average itself. Furthermore, on occasions the IMF may not be uniform across scale lengths of the order of the magnetosphere. The actual dynamical processes at the magnetopause - dayside merging, plasma entry - and the exact morphology of the magnetopause are topics of active current research.

Despite the considerable uncertainties involved, it is clear that the coupling between the solar wind and the magnetosphere results in the transfer and storage of energy in the Earth's magnetosphere. It is thought that this storage is in the form of reconfiguration of the Earth's magnetotail. Indeed, inflation of the magnetotail is often observed before substorm onsets. Interplanetary shocks can cause sudden compressions of the Earth's magnetosphere and can simultaneously excite or trigger substorm activity. In any event, energy can be stored in the Earth's magnetosphere and then be released later in a sudden manner.

Whereas the initial storage of energy may take place over many hours; and in fact apparently does so continuously albeit with varying efficiencies, the consequent energy dissipation in substorms takes place on time scales of an half hour. During substorms particles can be accelerated to high energies - 10's of KeV to  $\sim 1$  MeV. These particles are injected into the outer radiation zones where they subsequently drift around the earth. The electrons and ions are pitch angle scattered and eventually precipitate into the Earth's atmosphere causing various changes to the ionosphere which are of practical importance as noted in the Introduction.

#### Particle Precipitation at Low and Mid Latitudes

Ion precipitation at low latitudes has been observed to yield ionospheric effects. Qualitatively this may be understood in terms of charge exchange of ring current ions. As these ions become neutralized some disappear out into space and some precipitate into the near equatorial atmosphere. When they



encounter the upper atmosphere these energetic neutrals can become reionized and form a temporary low altitude trapped radiation zone from which further atmospheric interactions take place.

At low mid latitudes significant fluxes of energetic electrons are almost continually precipitating into the atmosphere as evidenced by the regular observation of quasi-trapped electrons at low altitudes and at longitudes where the minimum longitude drift altitudes are below the surface of the Earth. The precipitating fluxes at low latitudes are generally much lower than at high latitudes. Nonetheless, the fluxes can be detrimental to satellite missions and can interfere with communications.

Although middle latitude electron precipitation occurs most all of the time, large variations in flux are often observed from one spacecraft orbit to another. Thus substantial changes in the precipitation mechanism takes place on timescales shorter than  $\approx 100$  minutes. At the present these variations are only understood to a limited extent. Variations associated with magnetic storms are understood in terms of enhancement in the ambient radiation belt flux levels and the intensification of the scattering plasmaspheric hiss responsible for the post-storm precipitation. During other geomagnetic conditions our current knowledge is more meager. Enhanced ion convection across the plasmopause is thought to excite hiss generation and consequently variations in energetic electron precipitation at middle latitudes. Empirically, the precipitation fluxes show some increase with increasing Kp index, but this dependence is small in relation to the rather large spread in flux values observed.

Man-made VLF waves can propagate through the ionosphere and enter the radiation belt region. It has been suggested that these waves may be important in inducing energetic electron precipitation; but evidence in support of this has not yet demonstrated that this is a significant mechanism.

Our ability to predict short term variations in precipitating electron fluxes at times other than magnetic storm recovery is thus quite limited. Without a more complete understanding of the precipitation processes, most notably the generation of the scattering wave turbulence, the most appropriate approach to predicting short term variations may be to rely on the empirical finding that the total intensities of electrons precipitating from the slot region have significant correlation coefficients for separation times of 24 hours or less. Accordingly, without further research, only the persistence prediction appears useful: that the precipitating fluxes in the near future will be comparable to their present intensities.

#### Particle Precipitation at High Latitude

At high latitudes in the auroral zone/outer belt regions the fluxes of precipitating electrons are generally much greater than at lower latitudes and their morphology is more closely related to geomagnetic activity and to the occurrences of substorms. Since time delays (up to several hours) in the daytime precipitating fluxes may be associated with gradient drift, the known substorm morphology can presently provide a basis for predicting the fluxes of precipitating relativistic electrons in the dawn-noon time frame. The predictability of electron precipitation is complicated by the occurrences of



impulsive precipitation spikes at high latitudes, many of which occur in the midnight sector. To improve our predictive ability widespread precipitation patterns must be measured and for this purpose the satellite x-ray mapping technique can provide valuable data. With such a technique, from a polar orbiting satellite the spatial and temporal morphology can be studied in far more detail than has been heretofore possible.

### Prediction Techniques

The diamonds in Figure 1 represent the points at which observations might be made for predictive purposes. We shall briefly mention the types of phenomena which can be observed and the techniques which might be used to make further predictions.

(A) Solar flares may be observed in optical, radio and X-rays. Using models of typical proton events (see review by Shea and Smart in these Proceedings), predictions can be made for the appearance and duration of solar particle events at Earth.

(B) Observations can be made of the fluxes of energetic particles by satellites in interplanetary space. Such observations can be used to update model predictions.

Observations of energetic particles by spacecraft in magnetospheric regions easily reached by these particles (e.g., geosynchronous for protons of energies  $> 20$  MeV, electrons at the polar caps) can be used to predict entry of high energy protons into the tail lobes and precipitation onto the polar caps. Optical and X-ray imaging of the auroral region could be used to follow the course of a solar particle event as well as substorm evolution (cf. F and H below).

(C) Solar wind parameters (density, temperature, bulk velocity, components of the embedded magnetic field) can be measured upstream from the Earth. Various combinations of these parameters could be combined to give an empirical prediction of some index (e.g., Dst or AE) which is indicative of storm or substorm activity.

(D) Electron anisotropies obtained by spacecraft measurements near the geosynchronous orbit can indicate the occurrence of magnetic stresses in the magnetosphere. Direct observation of inflation of the magnetotail using magnetometer techniques can also indicate the existence of large amounts of stored energy in the magnetotail which might be released in the form of a substorm.

(E,F) Real time measurements of magnetic disturbances at ground stations can be used to track substorms. Particle injections can be observed by multiple spacecraft. Also multiple spacecraft observations of the magnetic field or symmetry axis for energetic electrons could be used to infer the parameters in a simple magnetic field model for the Earth's magnetospheric magnetic field. Theoretical solutions or computer modeling of the evolution of the observed particle distributions in the model magnetic field could be used to predict the motions of these particles and their eventual precipitation into the atmosphere.

(G) Spacecraft near geostationary orbit or in the geomagnetic tail can observe boundary motions or thinning on time scales of minutes prior to flux injections. For spacecraft in geostationary orbit, changes either in the gross counting rates due to the gradient in the flux profile or direct observations of the gradient for high energy proton fluxes (east-west effect) can be used to infer the inward convection of a boundary. The basic three-dimensional problem of thinning and the shape of the injection region must be understood in more detail. Local vs global effects must also be investigated.

(H) Direct observation of energetic flux levels and their pitch angle distributions can be used to predict precipitation of energetic electrons into the ionosphere. Coupled with particle trajectory calculations, these predictions might be extended to regions beyond the foot of the field line passing through the satellite.

#### Input from Other Working Groups

Much more research needs to be done into the substorm process to predict when and under what conditions we may expect energetic particles to be generated and to establish the connection between the magnitude of the generated fluxes and the state of the magnetosphere and the interplanetary medium. Theoretical studies should be carried out to describe the transport of injected particles in a realistic model of magnetic and electric fields. Much work has already been done using relatively crude field models for low energy plasma. Investigations should be directed towards simple magnetic field tail-like models which exhibit tail-like features and for electrons and ions with energies of less than 100 keV. Data comparisons should be made between magnetospheric particle observations and those data sets that bear on the coupling between the magnetosphere and the ionosphere and the upper atmosphere.

#### Testing and Evaluation of Predictions

There are at least four methods to test and evaluate prediction techniques. Table I lists pro's and con's for each of these methods. The methods are listed here as though they might be stages in the evaluation of a prediction technique.

(1) Traditional Collaborations. The traditional method for generating new scientific concepts and establishing new technologies is to form small collaborations. Publication in journals with refereeing and stimulation of further articles results in a winnowing process that allows the best ideas to emerge.

(2) Enhanced Collaborations. We have seen examples of the NASA and the NSF supporting analysis of special periods. This should allow the maximum number of people to participate and may allow certain groups to interact which might not have come together otherwise. New computerized technology has been used to facilitate comparisons of such special data sets.

(3) Simulated Prediction Schemes. Special events and computerized techniques developed in processes like (2) above could be used to test particular techniques.



(4) Field Tests. Particular techniques could be used by space disturbance forecasting centers as trials. Results could be used to refine the techniques.

Table 1. Methods of Testing and Evaluating Predictions.

	Pro	Con
Tradition Collaboration	Well established, comfortable	Slow process, not all data comparisons may be covered
Enhanced Collaborations	Geared to particular goal, new computerized techniques	Hardware limited, requires much individual time, few special events which may not be the critical cases
Simulation	Relatively cheap	Hardware limited, lacks subtleties of field situation, may lack realism.
Field Tests	Allows operating forecast center personnel hands-on experience and interactions with researchers	Very manpower intensive if carried out in real time



## APPENDIX II

### Long Term Magnetospheric Particle Variations ( $1 \text{ Day} < T < \infty$ )

J. I. Vette, H. K. Hills, G. A. Paulikas  
D. A. Sawyer, E. G. Stassinopoulos, and M. J. Teague

#### Introduction

The practical importance of studying the variations of energetic magnetospheric particles with durations in excess of 1 day lies mainly in being able to predict, for example, the effect of damage to spacecraft systems or long-term effects on the earth's ionosphere. Since typical systems, be these spacecraft systems or communication/navigation networks, have lifetimes in the range of 5-10 years, the variations of the near-earth environment over the solar cycle become an important aspect of their design and operation. With significant improvements in technology occurring in the same time interval, it is expected that such systems will continue to be replaced on similar time scales over the next half century. The radiation and plasma environment affects some aspect of engineering design and influences the selection of the orbit for certain space missions. With the total budgets of all nations for spacecraft systems and instruments being in the (several) billion dollar range annually, measures costing a few tens of million dollars might be considered quite cost effective.

From the scientific standpoint interest in long term variations of the radiation in nearby space ranges from the considerations of the evolution of life on the planet, where geomagnetic field reversals on geological time scales may result in significant changes in the radiation environment, to the detailed understanding of magnetospheric processes which can be extended to peculiar astrophysical objects such as pulsars and quasars. The complexity of plasma processes in the solar-terrestrial environment has been summarized in Appendix I and it has been shown that most phenomena operate in the time domain between one minute and one day. Consequently, it is necessary for us to logically impose a low pass filter with a cutoff frequency around one (day)<sup>-1</sup> to address our assigned topic; this is not an easy task. However, we will attempt to do this by adopting the viewpoint that such variations are either adiabatic (as concerns the motion of trapped particles), or the statistical results of non-adiabatic processes such as acceleration and, radial and pitch angle diffusion.

#### Long Term Energetic Particle Processes

Appendix I contains a comprehensive flow diagram of solar-terrestrial processes with observation points necessary for predictions and, hence, for understanding the phenomena (Figure 1 of Appendix I). The time scales relevant to the consideration of long-term particle variations are also indicated there.

The recognition of the variation of solar activity from one solar cycle to the next and the consequent variability of geophysical effects is manifest by the difficulties which were encountered in predicting the date of Skylab re-entry well in advance. Although sunspot numbers dating back to Galileo in the

early 1600's are available (including such anomalous periods as the Maunder minimum), it is still not possible to accurately predict the level of solar activity in a given solar cycle although the general variations can be predicted. The investigation of the Maunder minimum by such techniques as <sup>14</sup>C dating is near the extreme upper limit of quantitative study of long term radiation variability. On a shorter time scale, for example, we know that the persistence of active solar regions, which exhibit some periodicity around 45 days should also be included since it is suggested that such periodicities may be present in the energetic electron fluxes obtained in geostationary orbits over the past decade. Twenty-seven day running averages for this data show approximately 8 peaks per year and detailed correlations should be performed.

Recent studies of the interplanetary magnetic field for the period 1963-1976 has shown that the generation of shock structures is more prevalent during active periods. The character of the IMF and of long lived solar wind streams shows variations in the range of years. Consequently particle propagation from the sun also can be affected by the changes. The characteristics of the magnetosphere are also clearly affected by the long-term changes in interplanetary conditions. In addition to large scale erosion events, which are apparent when a large interplanetary shock strikes the magnetopause, there is recent evidence for patchy reconnection between the interplanetary magnetic field and the terrestrial field at the magnetopause. Thus buildup of the magnetospheric energy may occur by global as well as local processes. Thus it also seems plausible that de-excitation of the magnetosphere can occur in the same way, i.e., particle energization may be produced both impulsively and quiescently. We speculate that the "fireball event" in the magnetotail may be evidence of localized, patchy de-excitation and expect the particles from such acceleration events to show up on the outer magnetosphere in a less conspicuous manner than the large scale substorm injection processes, which may result from a neutral point formation across large regions of the geomagnetic tail. If one adopts this point of view, then there is evidence for de-excitation over the period of 10's of days and the magnetosphere may be capable of maintaining a stressed geomagnetic tail for these periods. This approach may be useful in organizing our view of magnetospheric dynamics since it offers a way of viewing the observed outer zone time behavior that brings some order to a very complex pattern.

The energetic outer zone particles, particularly electrons, exhibit large impulsive changes following the injection of accelerated plasma sheet particles into the trapping region. Although there are, on occasion, rapid (less than 1 day) changes, the general pattern is one of exponential decay with time constants in the range from days to 10's of days. Super-imposed on this are smaller injection events, which if not distinguishable, tend to make one believe the decay behavior has been altered. Studies of the decay constants indicate that they do not vary from event to event at a given L value and energy and the decay seems to be consistent with (but not in complete agreement with present calculations) whistler mode wave-particle interactions that result in pitch angle diffusion of the particles down the field line to be lost in the atmosphere. There are periods of time when these outer zone particles show a leveling off of the fluxes for days to 10's of days at a time. There is no reason to assume that the loss mechanisms become inoperative at this time, so we suppose a constant injection is occurring in a non-impulsive way. This



could occur from the integrated injections from a number of fireball or small de-excitation events.

Studies of outer zone electrons have revealed that following the arrival of fast solar wind streams, the fluxes rise within 1-2 days but then decay in a non-predictable manner independent of solar wind conditions. We view this decay as further evidence that the magnetotail is de-exciting itself over the period of some 20-30 days in small scale injection events. The observations of electrons  $> 4$  MeV show variability on the time scale of a year or two at geostationary orbits. However, at lower energies, averages of fluxes from 6 months to a year are consistent with there being no changes in the electron population over these periods in most of the outer zone down to  $L=4.5$ . Although definitive long term studies of protons have not been conducted, there is nothing to indicate their long term picture is different than electrons. Short term variations of protons including injection and decay have been studied. In the region between  $L=2.0$  and  $4.5$ , there are differences in the electron multi-month averages that can be as high as a factor of 10. Although the present quantitative models for trapped electrons attribute these changes to solar cycle effects, it is not clear that the observed periodicities support this conclusion.

An examination of the increases in energetic electrons in the outer zone following magnetic substorms has shown on some occasions time constants that exceed a day, particularly for the lower altitudes. Although some unknown long term acceleration deep within the magnetosphere has sometimes been attributed to cause these effects, it is much more likely that one is seeing the adiabatic transport of previously accelerated particles; this transport is due to the variation of the ring current. In the inner zone, one always finds rises greater than one day which are due to radial diffusion of the particles injected into the outer trapping region.

Starfish and other nuclear detonations provided a large perturbation in inner zone and slot electron fluxes that has allowed very good determinations of the loss rates due to atmospheric scattering and other factors. These exponential decay times were observed to be energy dependent with the longest times being about 370 days around  $L=1.45$  and energies around 1.0 MeV. Present quantitative electron models of the inner zone no longer contain a Starfish component since this residue has decayed below natural levels.

In the region around  $L=2.0$  energetic protons are known to have exhibited rapid non-adiabatic changes in association with large magnetic substorms. The subsequent recoveries are by radial diffusion. These types of events seem to reflect a slight re-distribution of particles rather than the impulsive injection of new particles. For the existing proton models the variations due to redistributions or long term variations are small enough (factors of  $\sim 2$  in restricted regions of space) that they have been ignored and the models are static except at low altitudes. Here the calculated predictions in 1964 that energetic proton fluxes would vary with the solar cycle have been confirmed and enough low altitude measurements were taken in various time periods to provide a somewhat quantitative description. Since these changes are produced by density increases in the upper atmosphere which in turn are caused by solar EUV emission that is dependent on coronal hole formation, the predictive aspects of



this phenomena are not without difficulty.

Finally we should note that changes in the internal geomagnetic field can produce long term changes in the energetic particle fluxes. Current internal field models possess secular changes to the dipole moment that indicate significant changes in times approaching 1000 years. The adiabatic changes due to this effect have been calculated. Further, the well known dipole reversals that occur on geological time scales may result in the complete disappearance of the magnetospheric cavity and the solar wind interaction would occur with the ionosphere as it does at Mars.

### Prediction Techniques

Prediction techniques for energetic particles within the magnetosphere depend on two kinds of knowledge. The first is the spatial variation of the trapped fluxes as determined by performing averages over at least several months at each point available for observation. Because the magnetospheric cavity itself undulates, changes shape, and moves in toto with respect to any preselected coordinate system, the separation of time effects, the second type of necessary knowledge, from spatial effects is never an easy task. However imperfect the situation may be, one proceeds by establishing some quantitative description of particle fluxes that to first order represents the average situation that any spacecraft would encounter in its flight profile. In a real orbit the time variations seen in executing each revolution may be as much as five or six orders of magnitude even using these static or quasi-static models. Added to this picture as a modulating influence are the inherent time variations themselves. The rapid departures from this quasi-equilibrium, which are represented by all the phenomena discussed in Appendix I, are spoken to separately within the phenomenology of the event in question. This is somewhat similar to conditional probability in that once an event has occurred, one can address some average properties of the event. For time variations that are long with respect to the multi-month averages constructed to obtain a spatial distribution, their cause is generally attributed to influences outside the system itself. Consequently, effects which show changes from years to several hundred years are believed to be due to solar processes or to transport processes between the sun and the earth. Processes that range from thousand years to several million years are associated with the internal dynamics of the earth such as the currents that produce the internal field. The time domain between 1 and 60 days is perhaps the most fruitful one for the magnetospheric scientist to study since the loss processes of most of the particles and the small scale de-excitations of the cavity lie mainly within this range.

The improvement of predictive techniques in this area is dependent on two things. First, more simultaneous observations of particle measurements are needed to obtain better spatial distributions. Not only are more spacecraft needed to perform these measurements but methods for providing the data collected to some centralized data base so it can be used effectively for producing models is needed. The use of Coordinated Data Analysis Workshops might be a way of improving the data bases needed for particle modeling and also bring into the modeling arena some of the experimenters involved in the original measurements. Unless there are major improvements in the numbers and types of instruments flown for these purposes, any enhancement in the collection of such

data coupled with added emphasis on the need for such studies represent the only pragmatic approach for improving the situation.

Since the motion of the energetic particles is governed by the magnetic field, the improvement of these models to cover disturbed conditions and the relaxation back to quiet conditions will be important element in improving the understanding of particle transport and decay. It is mainly through the magnetic topology of the geomagnetic tail that one can determine the energy stored in the tail. Knowing this quantity as a function of time, even crudely, would be a necessary parameter to testing the concept that de-excitation of the cavity by the acceleration of particles in patchy regions contributes to energetic particles in the outer trapping region.

#### Inputs from Other Working Groups

It is apparent from the discussion given in the previous section that inputs from the Working Groups on Interplanetary-Magnetosphere Interactions and Geomagnetic Disturbance Predictions concerning magnetic field models, magnetic energy stored in the tail, magnetic monitoring measurements in the geostationary orbit and at certain ground-based stations are the most important elements to study the configuration of particles in an adiabatic sense, once the non-adiabatic effects have occurred in the magnetosphere. Projects to concentrate on the change of energetic particle populations during certain events, particularly substorms, are fostered by coordinated measurements, as has become highly developed during the International Magnetospheric Study. Predictions of possible magnetospheric events from solar events have also proved useful to satellite measurements in the past. If the data from ISEE 3 could be made available in real time to a prediction center, then alerts for interplanetary shock waves and other solar generated disturbances could be used to improve data coverage prior to and during the magnetospheric response to such events.

#### Testing and Evaluation of Predictions

The understanding of a process improves the ability to predict. Even a statistical understanding, as opposed to a physical one, provides a statistical prediction. The latter techniques are used very extensively in doing predictions on solar events since the complexity of the processes are indeed very great. The general methods of evaluation have been provided in Appendix I. The specific viewpoint we have taken in this position paper is certainly one that is subject to test. Can all of the long term decay processes displayed by magnetospheric particles be shown to be governed just by internal processes? The effect of magnetospheric events seems to result in a general increase of the particle population save for a few isolated regions where non-adiabatic transport might result in a catastrophic increase in loss rates. It is felt that enhanced collaborations, particularly through Coordinated Data Analysis Workshops as well as traditional methods will be the most fruitful in evaluating the new prediction techniques which may become available through an increased study of decay processes and their relationship to the energy content of the various regions of the magnetosphere.



# ENERGETIC SOLAR PARTICLE BEHAVIOUR IN THE MAGNETOSPHERE

M. Scholer  
Max-Planck Institut für Extraterrestrische Physik  
Garching b. München  
West Germany

The behaviour of energetic solar flare particles in the magnetosphere is discussed. In the absence of magnetospheric motion, the problem of particle transport can be treated as simple propagation of charged particles in a stationary magnetic field configuration using, for instance, trajectory calculations in model fields. This single particle approach is the basis for the determination of intensity and anisotropy structures over the polar caps and in the geomagnetic tail from different interplanetary conditions. Particle transport on closed fieldlines is in addition strongly affected by resonant interaction processes as pitch angle scattering and radial diffusion.

## 1. INTRODUCTION

Energetic solar flare particles, once they hit the magnetosphere, enter the geomagnetic field and can be observed over the polar caps, on closed field lines in the inner magnetosphere (e.g. at synchronous orbit) or on open field lines in the geomagnetic tail. Depending on the interplanetary or magnetosheath particle and magnetic field conditions, on the particle energy and on the observation position within the magnetosphere, a great variety of intensity-time structures, spatial intensity structures and anisotropies will be observed. This report tries to summarize the basic modes of particle propagation to the various observational regions, which are the link between the conditions in interplanetary space and the observed structures within the magnetosphere.

The motion of a particle in a homogeneous magnetic field is a spiral and the motion of the center of gyration is given by the field line to which the particle is tied. The magnetic field within the magnetosphere is strongly inhomogeneous and it is necessary to state the limits of accuracy of the guiding center approximation. Figure 1 shows curves of constant rigidity separating regions of adiabatic and nonadiabatic particle motion on a meridional cross-section of the magnetosphere (Morfill, 1973). The constant rigidity contours are defined by the condition that the gyroradius is one-tenth of the characteristic length's scale of the magnetic field. Since the



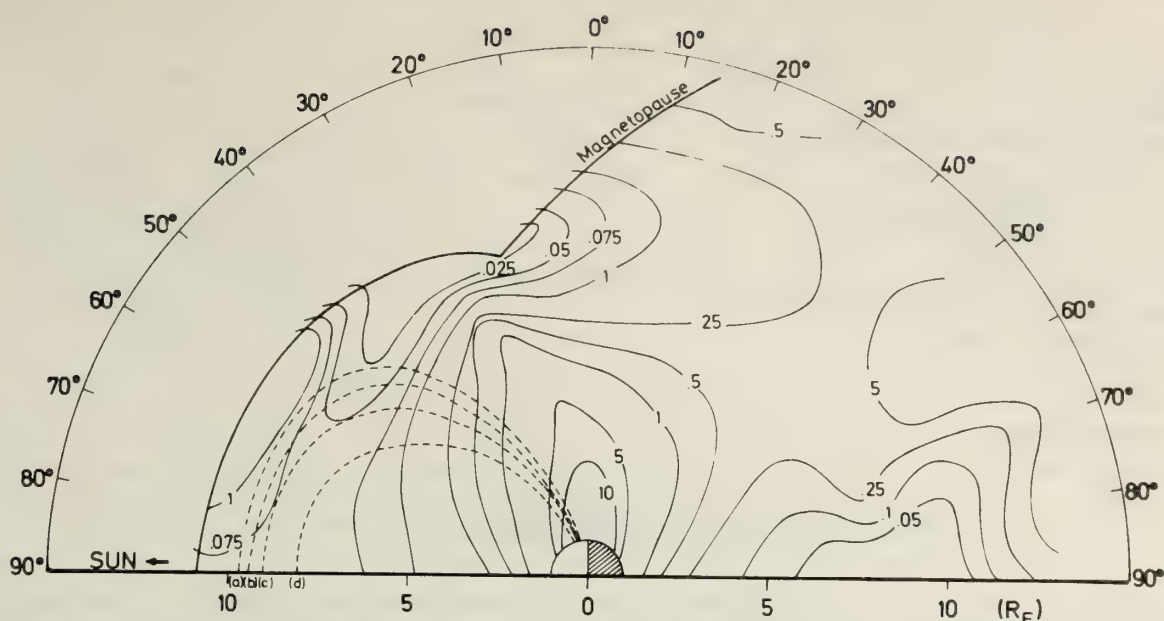


Figure 1. Curves of constant rigidity separating regions of adiabatic and non-adiabatic particle motion, as obtained from the Alfvén criterion  $\rho(\nabla \cdot \mathbf{B})/B = A = 0.1$ . The rigidity is calculated from  $R = 300 B^2 A / \nabla \cdot \mathbf{B}$ , and a cross-section through the magnetosphere in the X-Z plane is shown (Morfill, 1973).

iso-rigidity contours give the highest rigidity particles which move adiabatically, we see that there are large regions in the magnetosphere where even 1 MeV protons (0.04 GV) do not propagate under the adiabatic approximation. For all these particles, the full equation of motion must be solved in order to obtain a reasonable model description of their orbit in the magnetosphere. Further out, beyond  $\sim 14 R_E$  in this magnetic field model (Williams and Mead, 1965 field model), the geomagnetic tail consists essentially of parallel field lines, so that the propagation of particles with larger rigidities is again adiabatic in the tail. The nonadiabaticity implies that we have to numerically integrate particle orbits in a quantitative magnetospheric field model at least out to  $15 R_E$  or to the magnetopause in order to infer the particle distribution function close to the earth from interplanetary measurements or to draw conclusions from measurements of the distribution function close to the earth. The application of trajectory integrations assumes that the geomagnetic field is static; the effects of electric fields, when integrated over the trajectory, must be small and so must pitch angle scattering and cross  $\mathbf{B}$  diffusion. In other words, the distribution function  $f(\mathbf{p}, \mathbf{r})$  of the cosmic rays must be conserved along the trajectory. This is of course not always true and is especially violated within the closed field line region, as will be discussed later.

Another important quantity is the direction of asymptotic approach schematically shown in Figure 2 (from Shea et al., 1965) and refers to the direction from which the solar cosmic rays arrive before they enter the geomagnetic field. This ultimately tells us how the interplanetary anisotropy is transferred to the observation point within the magnetosphere.

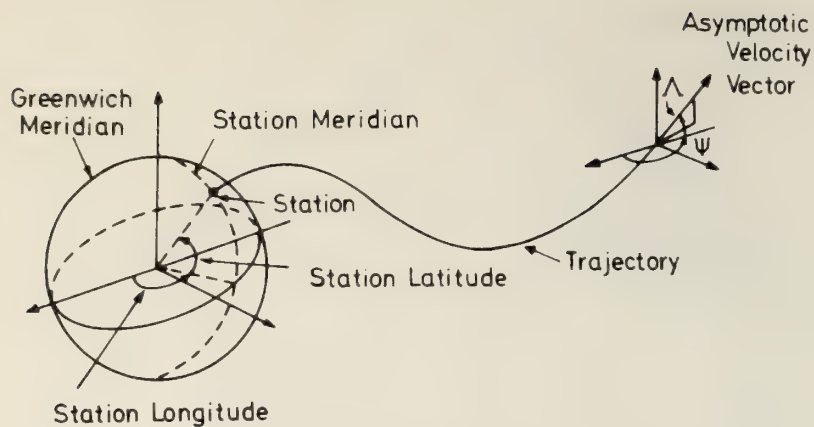


Figure 2. Illustration of the definition of the asymptotic direction of approach (Shea et al., 1965).

## 2. ENERGETIC SOLAR PROTONS OVER THE POLAR CAPS

Systematic trajectory calculations in a Mead-Williams (1965) field model have been performed by Morfill and Quenby (1971). Figures 3 and 4 show polar cap regions of vertically incident particles in which particle access occurs either from the dawn or dusk flank of the magnetopause or the high latitude magnetotail and the neutral sheet, respectively. Figure 3 shows the

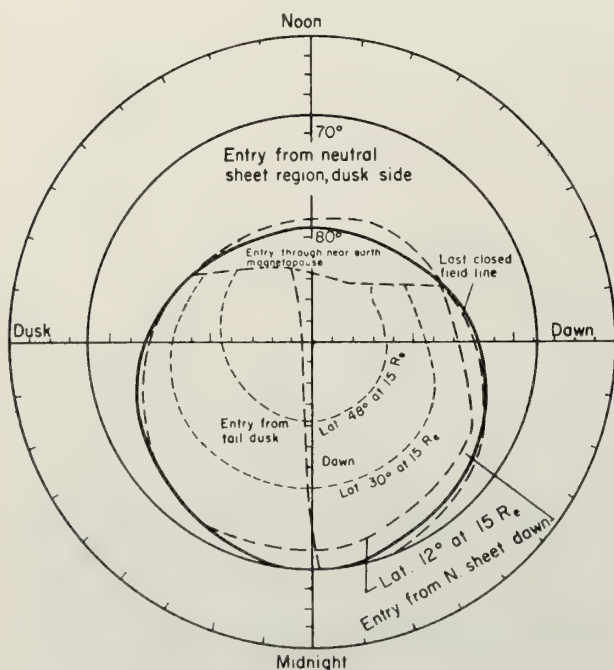


Figure 3. Trajectory computed access regions over the northern polar cap for solar protons, showing entry through the magnetopause, from the tail and the plasma-neutral sheet regions for 0.08 GV particles (Morfill and Quenby, 1971).

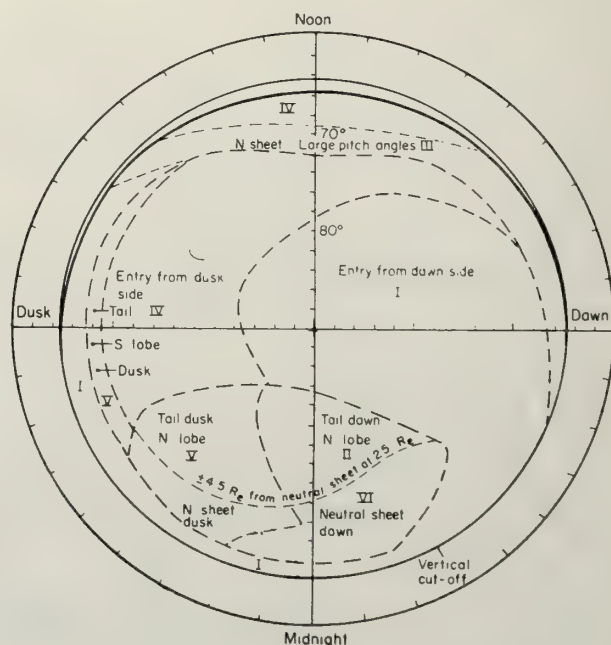


Figure 4. Same as Figure 3 for 0.5 GV particles.

resultant access regions for a 3 MeV proton and Figure 4 shows the access regions for a 125 MeV proton. Let us first discuss the higher energy particles. There is a large region over the central polar cap where access is by direct impact from either the dawn side or dusk side near earth magnetopause. Towards the midnight central polar cap the particle trajectories end in the central part of the northern tail lobe. Around the vertical cutoff is a broad region, where access is from the neutral sheet. At  $25 R_E$  the particle trajectories end either at the dawn or dusk side of the neutral sheet. Figure 5 shows schematic projections of trajectories onto the ecliptic plane of the magnetosphere starting at different points along a dawn-dusk satellite pass over the polar cap. Particles observed in a small range of latitudes at  $\sim 64^\circ$  invariant latitude at dawn and dusk enter the magnetosphere on the dawn side by direct impact and arrive from a direction close to the interplanetary spiral angle. Particles at somewhat higher latitudes have trajectories close to the neutral sheet which end up by drifting out on the dawn side of the magnetopause. At somewhat higher latitudes at dusk, particles enter on the dusk side of the magnetopause by direct impact and at still higher latitudes, trajectories end up in the tail. High energy protons usually exhibit in interplanetary space an anisotropy in the Parker spiral direction. The direct impact from the dawn and dusk side results then in auroral peaks at both dawn and dusk and in a trough at somewhat higher latitudes near dusk.

In the case of an interplanetary anisotropy, the full intensity is usually also observed over the central polar cap where trajectories end in the central tail. Two mechanisms have been proposed in order to explain the high polar cap enhancements at large energies: In the reconnection model, which implies divergence of the tail field, all field lines, even those from the center of the tail, are eventually brought to the surface. By considering an idealized rotational discontinuity for the interface and by calculating trajectories in such a static field configuration, Durney et al. (1972) were able to show that multiple discontinuity traversals of a particle produce large angle changes in particle velocity vector, so that particles coming from the sun could eventually be turned around at the interface by more than  $90^\circ$  (Figure 6). A different explanation has been given by Willis et al. (1973). Willis and Pratt (1972) constructed a model of the tail field where all field lines connect through the neutral sheet. Whereas in the Mead-Williams field model trajectories end at  $25 R_E$  in the central tail,

#### ECLIPTIC PLANE CROSS SECTION

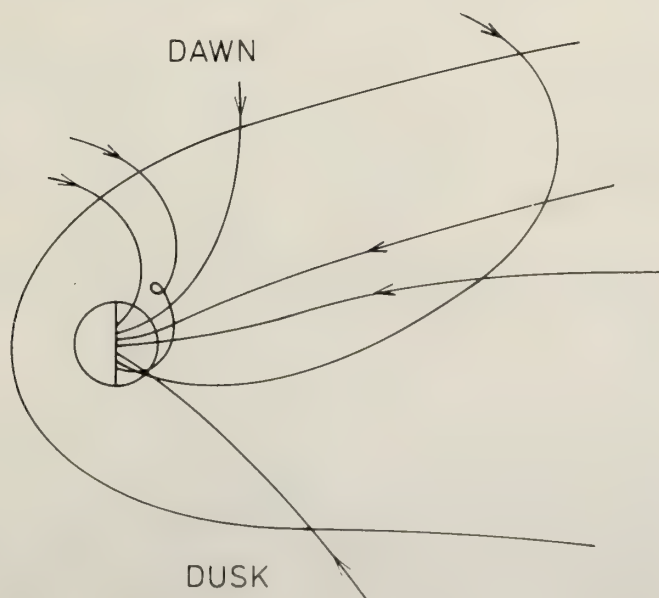


Figure 5. Schematic projections of trajectories of 130 MeV protons onto the ecliptic plane of the magnetosphere.



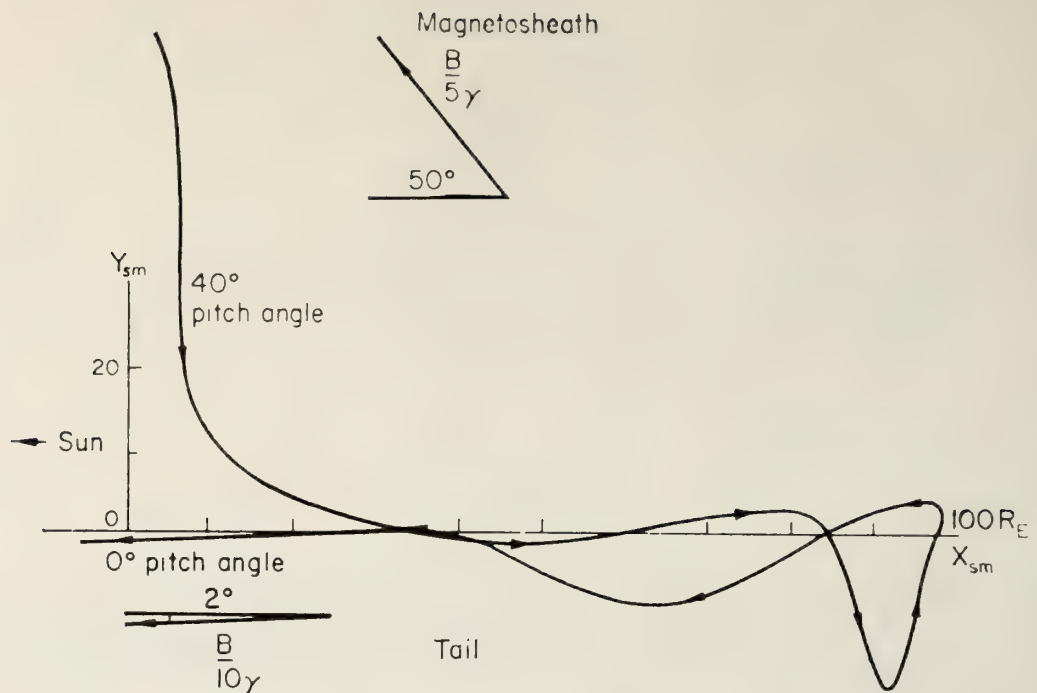


Figure 6. Proton motion across a discontinuity for multiple discontinuity traversal produce large angle changes in particle velocity vector. Final pitch-angle in the tail is  $0^\circ$  in this example (Durney et al., 1972).

in the Willis-Pratt model all trajectories end ultimately near the neutral sheet and are turned by the residual  $B_z$  component towards dawn, so that the full intensity is accessible. Figure 7 shows the magnetopause entry diagram of a dawn-dusk polar orbit; i.e. the termination of particle trajectory integrations are shown on a latitude-longitude solar-magnetospheric coordinate system. It can be seen that in the Mead-Williams model the central polar cap is at  $25 R_E$  still connected to the central tail, whereas in the Willis-Pratt model these trajectories have crossed the magnetopause near the dawn neutral sheet before  $100 R_E$ .

Coming to the lower energy protons, it can be seen from Figure 3 that the polar cap region accessible by direct impact becomes smaller and smaller and is located near the last closed field line at noon. Above the last closed field line, entry is from the tail; in the so-called pseudotrapping region, entry is from the neutral sheet. In the pseudotrapping region, particles must drift in longitude (protons from dusk to dawn via noon) in order to arrive at a specific local time; thus they enter from the dusk side of the neutral sheet. However, as has been stated above, the residual  $B_z$  component of the magnetic field results in a turning for particle trajectories crossing the neutral sheet, so that all proton trajectories integrated backwards from the earth and crossing the neutral sheet finish up on the dawn side of the magnetopause. Figure 8 shows a 20 MeV proton trajectory with multiple neutral sheet crossings (Morfill and Quenby, 1971). As can be seen, the bending of the particle trajectory produces access to the dusk auroral zone from the Parker spiral direction. The length of trajectory in the tail is usually less than 100 earth radii for small initial pitch angles, i.e. the protons have entered the tail close to the earth. This is verified

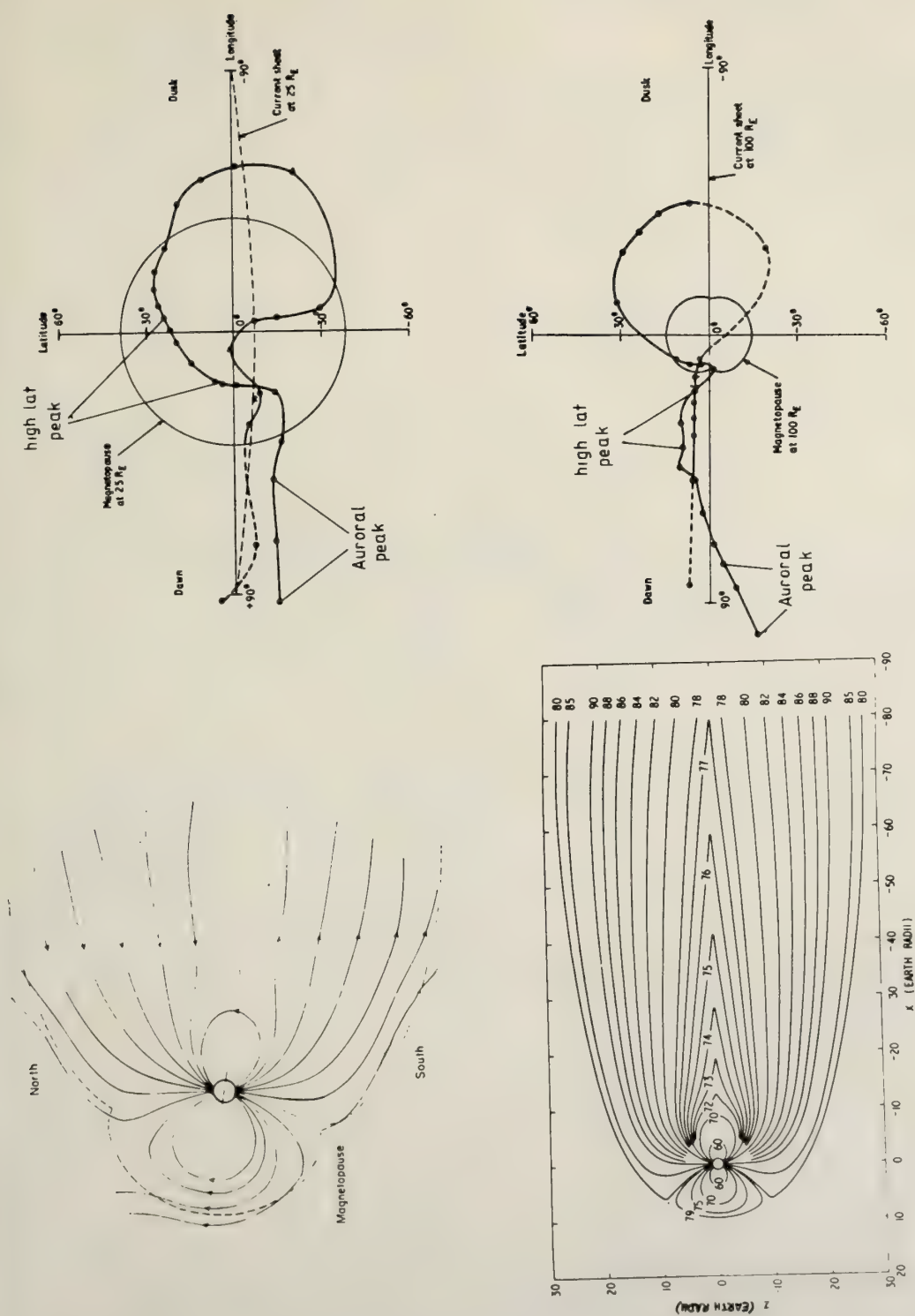


Figure 7. Comparison of trajectory calculations in the Mead-Williams field model and the Willis-Pratt field model. Shown are the positions of 125 MeV protons starting on a dawn-dusk polar cap orbit at the termination of particle trajectory integrations.

experimentally; Figure 9 shows the location of solar proton access windows in the magnetotail as a function of proton energy, as measured by various workers (Paulikas, 1974).

Another interesting aspect of the turning mechanism of low energy particle trajectories crossing the neutral sheet is the dispersion of trajectories with different initial pitch angles. Trajectories starting at the Vela 4A position in the tail with different initial pitch angles are shown in Figure 10 (Morfill and Quenby, 1971). Trajectories starting near the neutral sheet with large initial pitch angles leave the tail closer to the earth than those with small initial pitch angles and one has to be careful when deriving access windows or tail lengths from such measurements.

Coming back to low energy protons over the polar caps, we have seen that the central polar cap is connected to the central tail lobe. In a reconnection model of the magnetotail, these field lines ultimately end up at the magnetopause far down the tail where they connect to the interplanetary field. For high energy protons, the magnetopause will look like a tangential discontinuity as discussed above; lower energy protons leave a correspondingly smaller gyroradius (e.g. a typical gyroradius of a 1 MeV proton at  $100 R_E$  from the earth is  $\sim 4 R_E$ ) so that these particles could follow reconnected field lines adiabatically. This is not borne out by measurements; it seems that there is considerable scattering of  $\sim 1$  MeV protons near the magnetopause, possibly due to a turbulent boundary layer. Figure 11 from a recent paper by Palmer et al. (1979) shows in the upper panel interplanetary pitch angle distributions. These distributions are then convoluted with a Gaussian scattering function ( $\sigma = 0.3$ ) to give the distribution transmitted through the magnetopause. The dashed curve is

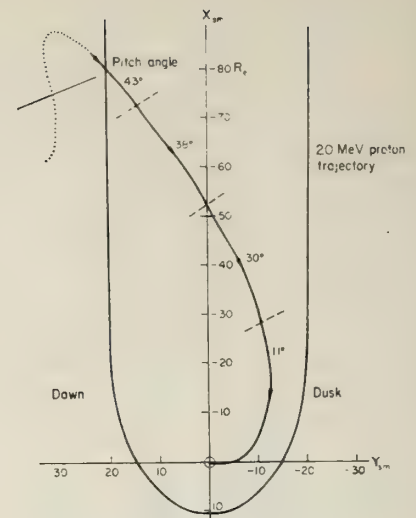


Figure 8. 20 MeV proton trajectory with multiple neutral sheet crossings, showing the change in particle pitch angle after each crossing, and the bending of the particle trajectory which produces access to the dusk auroral zone from the Parker spiral direction (Morfill and Quenby, 1971).

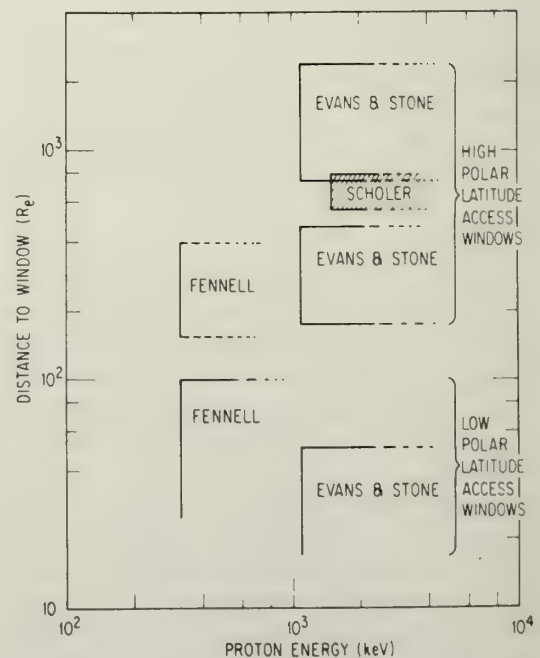


Figure 9. Location of solar proton access windows in the magnetotail as a function of proton energy, as measured by various workers (Paulikas, 1974).



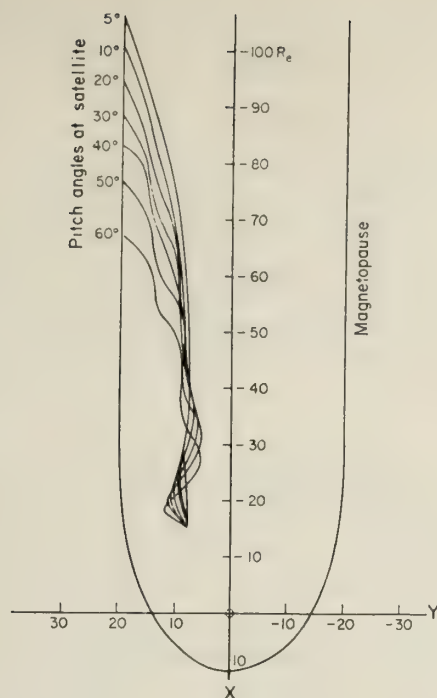


Figure 10. Dispersion of low energy proton trajectories, in a realistic tail field model, for the Vela 4A position at 14.30 UT during the September 18th, 1967 flare. Dispersion is due to pitch angle variations and the multiple neutral sheet crossings involved (Morfill and Quenby, 1971).

the convoluted distribution transformed to the satellite position, assuming a threefold increase in magnetic field strength between entry point position and satellite position and corresponds well with the measured distribution. The scattering strength at the magnetopause seems variable; other authors have reported total isotropisation within the tail (e.g. Montgomery and Singer, 1969).

We now summarize briefly solar particle intensity structure to be observed in the open field line region. Figure 12 (Scholer, 1975a) is a schematic representation of different interplanetary particle conditions leading to

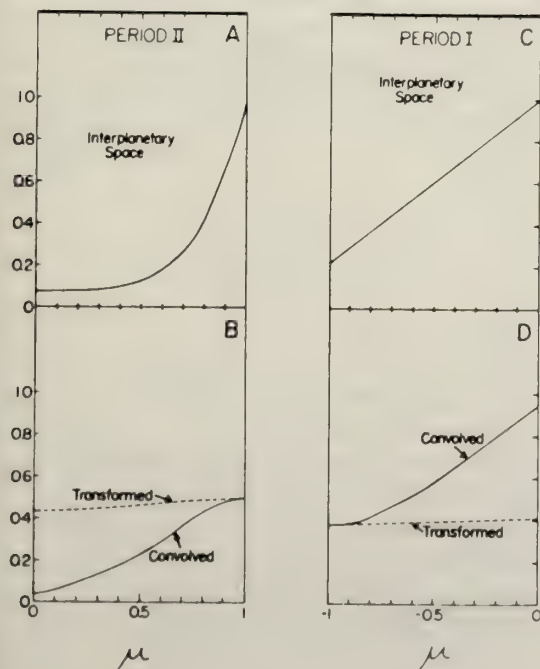


Figure 11. (a) Interplanetary pitch distribution incident on magnetopause above the north lobe. (b) Convolution of this distribution with a Gaussian scattering function of  $\sigma = 0.3$ , to give the distribution transmitted through the magnetopause. The dashed curve is the convolution transformed to  $r \approx 18 R_E$ , a threefold increase in magnetic field strength being assumed. (c) and (d) same for another time period (Palmer et al., 1979).

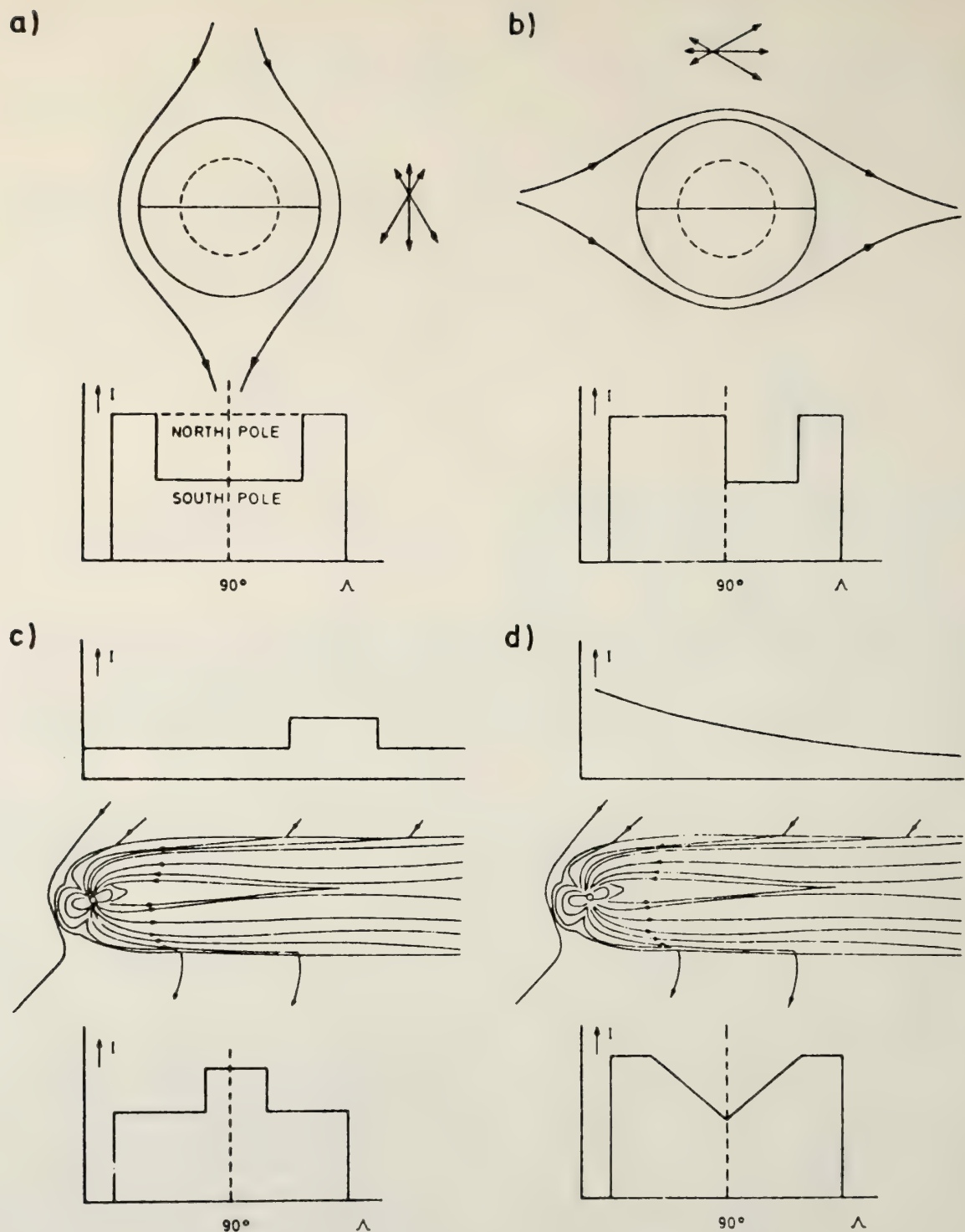


Figure 12. Schematic representation of different interplanetary flux conditions leading to polar cap structures: (a) interplanetary north-south anisotropy leading to a north-south polar cap asymmetry, (b) interplanetary east-west anisotropy leading to a dawn-dusk asymmetry over both polar caps, (c) convected filament with higher flux leading to "hump" over the polar cap, (d) interplanetary intensity gradient leading to corresponding gradient over the polar cap.

correspondingly different structures over the polar cap. Figure 12a shows in the upper part a cross-section of the tail with some field lines draped around in the north-south direction.

An interplanetary field aligned anisotropy will illuminate the north tail lobe with the higher flux and the south lobe with the lower flux. A dawn-dusk polar orbiting satellite will probe the tail along the dashed line and consequently a north-south asymmetry will result (with equal intensities in the closed field line region). In the case of an interplanetary field in the ecliptic with an east-west anisotropy (Figure 12b) isotropisation of the particles at the magnetopause should lead to an east-west asymmetry over the polar caps, as has indeed been observed by e.g. Haskell and Hynds (1972) and Hynds et al. (1974).

Tangential discontinuities in interplanetary space can separate regions of different proton flux. This is indicated in Figure 12c where we have assumed that a magnetic filament, containing a higher flux than the surrounding region, has been convected along the magnetotail with solar wind speed. This will lead to a corresponding flux increase over the polar cap on field lines which leave the magnetopause at the position of the filament. Observations of such structures have been particularly important in order to measure access windows along the tail (e.g. Evans, 1972; Scholer, 1972; Fennell, 1973; Hynds et al., 1974). Finally, during the first stage of solar particle events, large negative gradients can exist in interplanetary space. For a tail length of  $\sim 0.1$  AU, this can lead to a 20 - 40% intensity difference along the tail as schematically indicated in Figure 12d. Thus, a gradient between the low polar cap region, where asymptotic directions go out close to the earth, and deep in the polar cap, where field lines go right to the end of the tail, can be possible.

To end the discussion of structures in the open field line region, Figure 13 shows orbits of two polar satellites projected into the geomagnetic tail during the passage of an interplanetary discontinuity behind of which the flux was decreased (Hynds et al., 1974). It can be seen that a tube of higher particle flux exists in the center of the tail indicating that this field line crosses the magnetopause way down the tail at a point which the discontinuity has not yet reached. This is schematically indicated in Figure 14.

### 3. ENERGETIC PROTONS ON CLOSED MAGNETOSPHERIC FIELD LINES

#### 3.1 Quiet Magnetosphere

The last part of this review deals with solar particle behaviour on closed magnetospheric field lines. We have already seen that high energy protons can reach the closed field line regime by direct impact, whereas lower energy protons enter the pseudotrapping region from the neutral sheet. Due to the nonadiabatic behaviour of energetic particles in the inner magnetosphere, solar flare protons can also reach synchronous altitude. Trajectory calculations for particles arriving at synchronous orbit (Gall et al. 1971) show that particles near the geomagnetic cut-off have access at any local



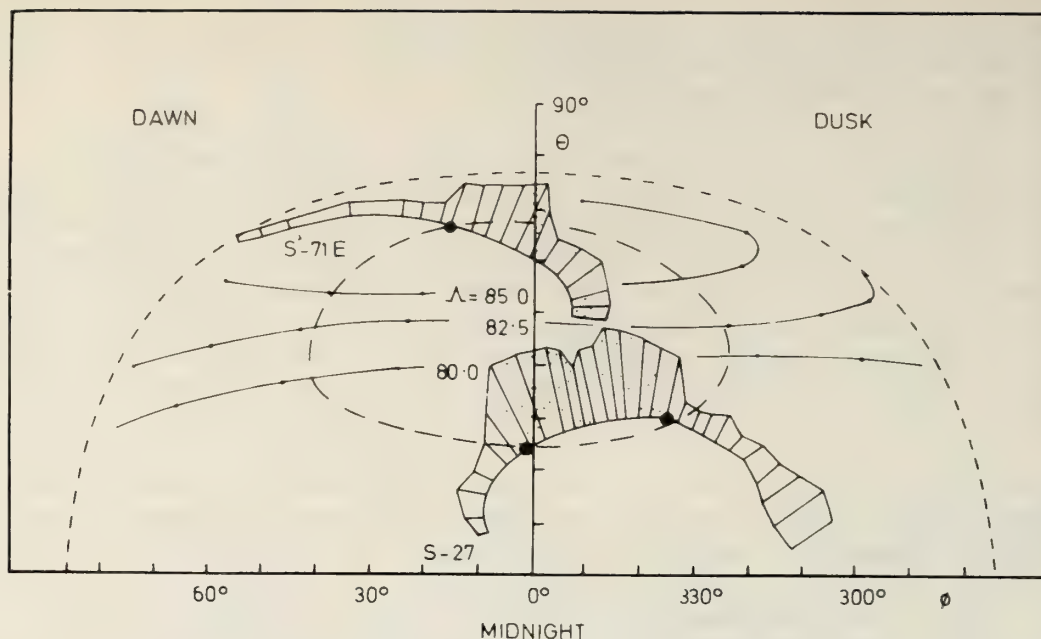


Figure 13. Projections of satellite orbits (ESRO 1, S-71E and ESRO 2, S-27) along 0.08 GV trajectories to 15 Earth radii in the tail during the November, 1968 flare, after the passage of an interplanetary tangential discontinuity past the geomagnetic cavity. The measured particle intensities are represented as the lengths of the lines drawn normal to the orbit projections (Hynds et al., 1974).

time from the tail after having drifted once or twice around the earth. The cut-off rigidity, i.e. the lowest rigidity so that a particle still can reach the synchronous orbit, is strongly dependent on the angle of incidence at the synchronous orbit. Particles arriving from west have easiest access and therefore this direction has the lowest cut-off rigidity. Figure 15 shows the cut-off energy for protons arriving at ATS 1 vertically and from local west vs. local time. As can be seen, the cut-off energy is strongly local time dependent and at 24 h LT protons with energies as low as  $\sim 3$  MeV can reach the satellite via a direct trajectory. The difference in cut-off

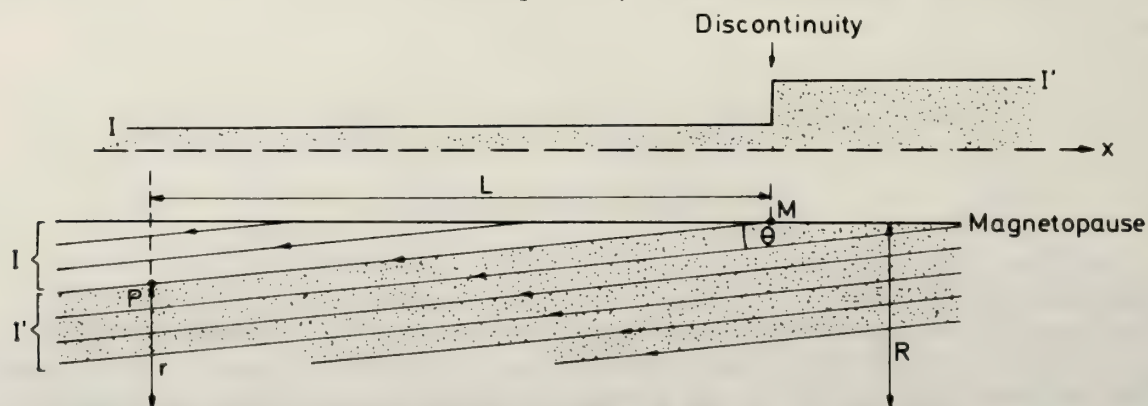


Figure 14. Schematic cross-section through one tail lobe in order to demonstrate the effect of a discontinuity in the particle intensity caused by interplanetary magnetic field structure.

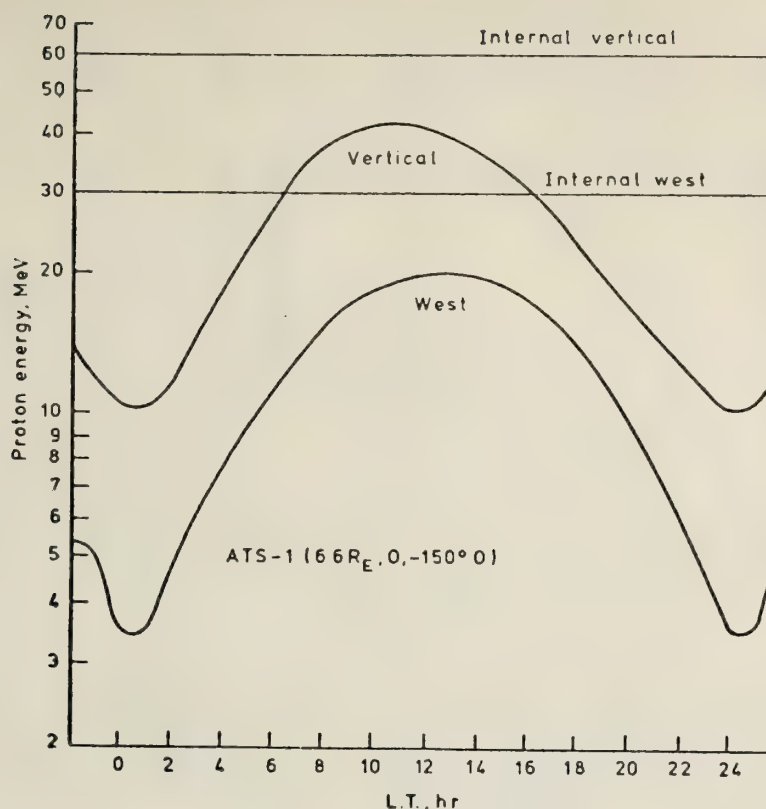


Figure 15. Daily variations of the computed cut-off for protons arriving at the ATS 1 geostationary satellite vertically and from local west. Cut-offs corresponding to the local time independent internal field are also shown (Gall et al., 1971).

energies for incidence from the east-west or vertical direction will lead to local anisotropies at the synchronous altitude.

Blake and Friesen (1977) have recently calculated access cones for energetic particles reaching synchronous orbit from the near earth magnetopause. Their results are shown in Figure 16. The access cone, i.e. the angular sector centered approximately around the equator wherein particles coming from the near earth magnetopause can be observed, is strongly energy and local time dependent. The access shrinks to zero at  $\sim 17$  MeV (at 0600 LT); no access occurs for 30 MeV protons later than 0800 LT. This demonstrates that determination of access to synchronous orbit is extremely complicated and depends critically on local time, instrument look direction and energy.

We have seen that low energy ( $\sim 1$  MeV) particles in the pseudotrapping region enter near midnight from the neutral sheet and drift around the earth. The last closed drift shell is given by the last closed field line at midnight, as shown in Figure 17 (Morfill, 1973). Also indicated in Figure 17 is the region (hatched area) in the pseudotrapping region where breakdown of adiabatic motion occurs due to the field line curvature near the last closed field line. Breakdown of adiabatic motion leads to the effect that the trajectory derived loss cone at the equator is different from the adiabatic loss cone, or in other words, a particle starting at the trajectory derived

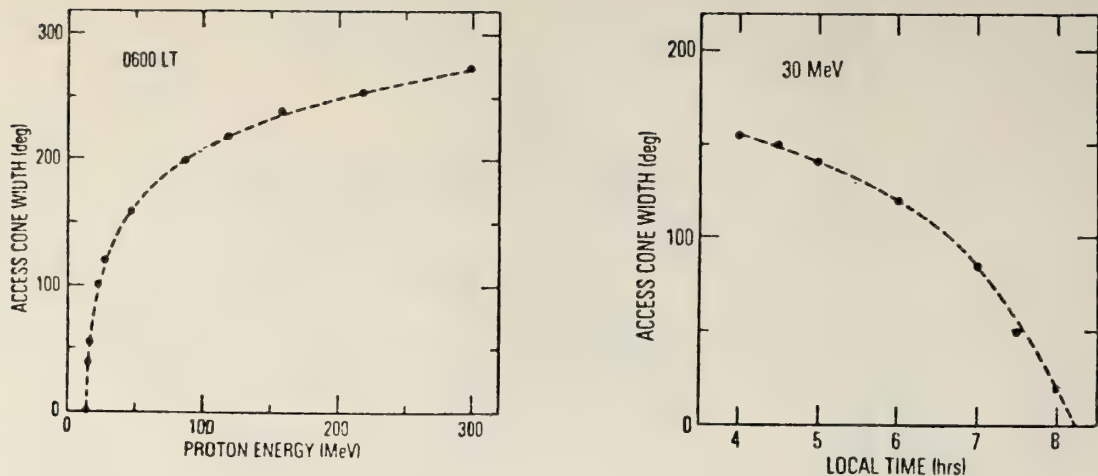


Figure 16. (a) Access cone width as a function of proton energy.  
 (b) Access cone width as a function of local time.  
 (Blake and Friesen, 1977).

loss cone at the equator will not be observed in the opposite hemisphere at 100 km with  $90^\circ$  pitch angle. Thus at low altitudes the upper loss cone will be filled with particles as in the case of strong pitch angle scattering. This is of course the same as saying that the trajectories in this region of the magnetosphere are not reentrant.

The extent of the nonadiabatic region near midnight defines the last adiabatic drift shell; between the last closed drift shell and the last adiabatic drift shell, particles that drift completely around the earth undergo changes in pitch angles and orbits, owing to breakdown of the adiabatic invariants each time they enter a localized region near midnight. The same process allows access of solar protons down to the last adiabatic drift shell and explains the lowering of particle cutoffs even at quiet times to below the computed cutoff rigidities. Particle trajectories starting in this region circle several times the earth before they enter the neutral sheet region. This has been demonstrated by the fact that after sudden flux decreases in interplanetary space, the flux of solar particles at latitudes below the auroral region still stays high for quite some time (Scholer, 1975b), i.e. solar particles drift several times before they are finally lost again by the accumulative effect of nonadiabatic motion when they traverse the region close to midnight.

### 3.2 Dynamic Effects of the Magnetosphere

So far, we have discussed the effects of the static magnetospheric field configuration on the particle motion. Clearly, there can be additional dynamic effects of the magnetosphere on solar particles. One of these dynamic effects is pitch angle scattering of the particles during their drift from dusk to dawn in the pseudotrapping region and in the stable trapping region. If pitch angle scattering is strong, the atmospheric loss cones are filled and emptied during each bounce period so that the equatorial particle



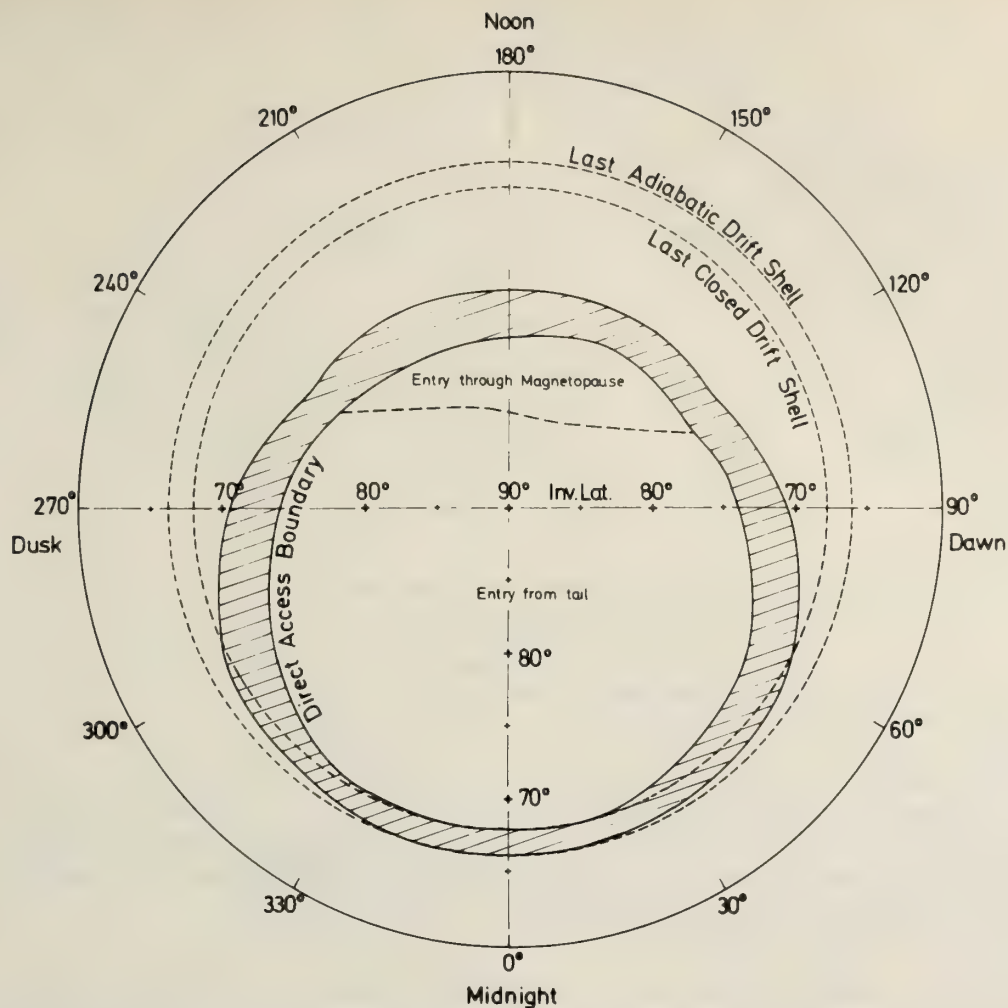


Figure 17. Polar diagram of access regions, quasi-trapping regions and positions where breakdown of adiabatic motion occurs. The direct access boundary and the lower boundary where breakdown of the guiding center approximation occurs are drawn solidly, the hatched area representing that part of the pseudo-trapping region where nonadiabatic particle motion in a spatially inhomogeneous field does not allow double loss cone formation (Morfill, 1973).

distribution is isotropic or the upper loss cone at low altitudes is filled. Such single loss cones and isotropic distributions have been observed in the dusk part of the nightside during the main phase of a magnetic storm. At the same time, the particles at the day side show an anisotropic loss cone distribution. During the recovery phase of the storm, the dayside anisotropy vanishes and the proton fluxes are isotropic through the loss cone.

Figure 18 (Williams and Heuring, 1973) shows schematically the evolution of a low-altitude anisotropic loss cone distribution at three stages of a geomagnetic storm. The disappearance of the anisotropic pitch angle distribution can be interpreted as the result of strong pitch angle scattering associated with the development of a symmetric ring current. In the recovery phase of the storm, the ring current particle population is symmetric and spatially envelopes the solar proton distribution also on the dayside hemisphere. At

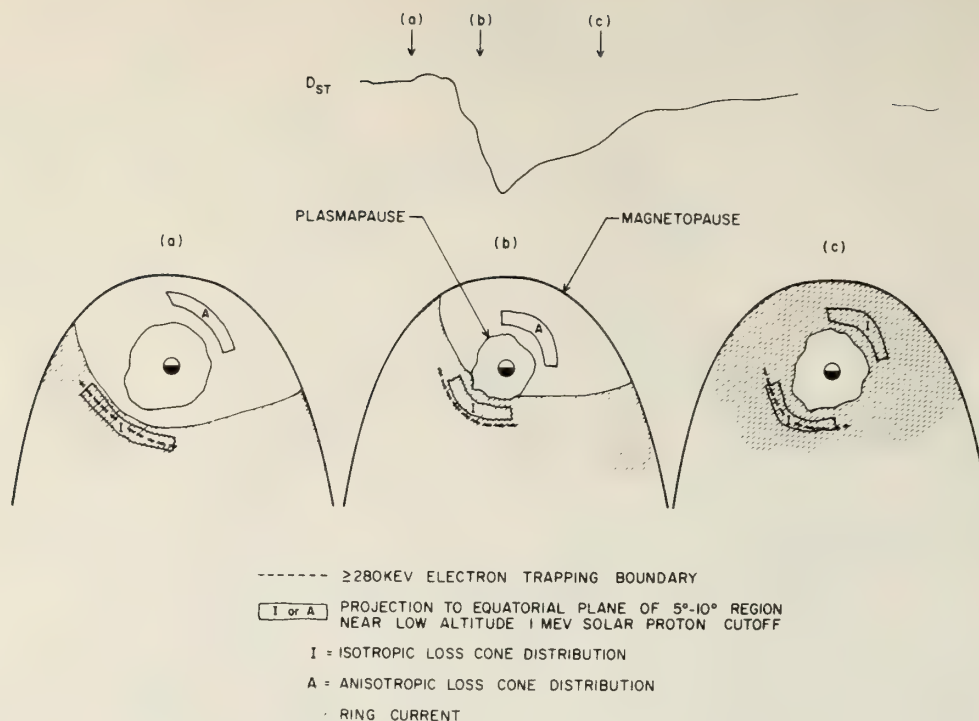


Figure 18. Schematic of the evolution of a low-altitude anisotropic proton loss cone distribution into an isotropic loss cone distribution. Three time periods in a typical Dst magnetic storm are shown. Dayside anisotropy is seen before the storm and during the asymmetric (main) phase (Williams and Heuring, 1973).

quiet times, the power in the magnetic fluctuations at large frequencies is too low, so that normally one would not expect strong pitch angle scattering of energetic protons in the magnetosphere to occur. However, it was argued by Scholer and Morfill (1975) that during periods of strong ring current the additional ring current particles reduce the Alfvén velocity, which shifts the resonance interaction frequency towards lower values. If the wave power spectrum is steep enough, more power is now available at the resonance frequency and the condition for strong pitch angle scattering can be satisfied.

In the case of weak pitch angle scattering, the omnidirectional equatorial intensity remains approximately constant; the main effect is to change the equatorial distribution to a distribution peaked at  $90^\circ$  by losing particles near the nominal loss cone to the atmosphere. This is schematically shown in Figure 19. A low altitude satellite samples only a small range of equatorial pitch angles; thus in a region of weak scattering, the flux at low altitudes can decrease substantially even though the total number of particles in the flux tube decreases by only a small fraction.

Disturbed geomagnetic conditions, as during geomagnetic storms and substorms, will also clearly influence transport characteristics of solar particles in the closed field line regions. The equatorial shift of the cut-off latitude observed during the main phase of a geomagnetic storm can be caused by the presence of a symmetric ring current. Figure 20 shows

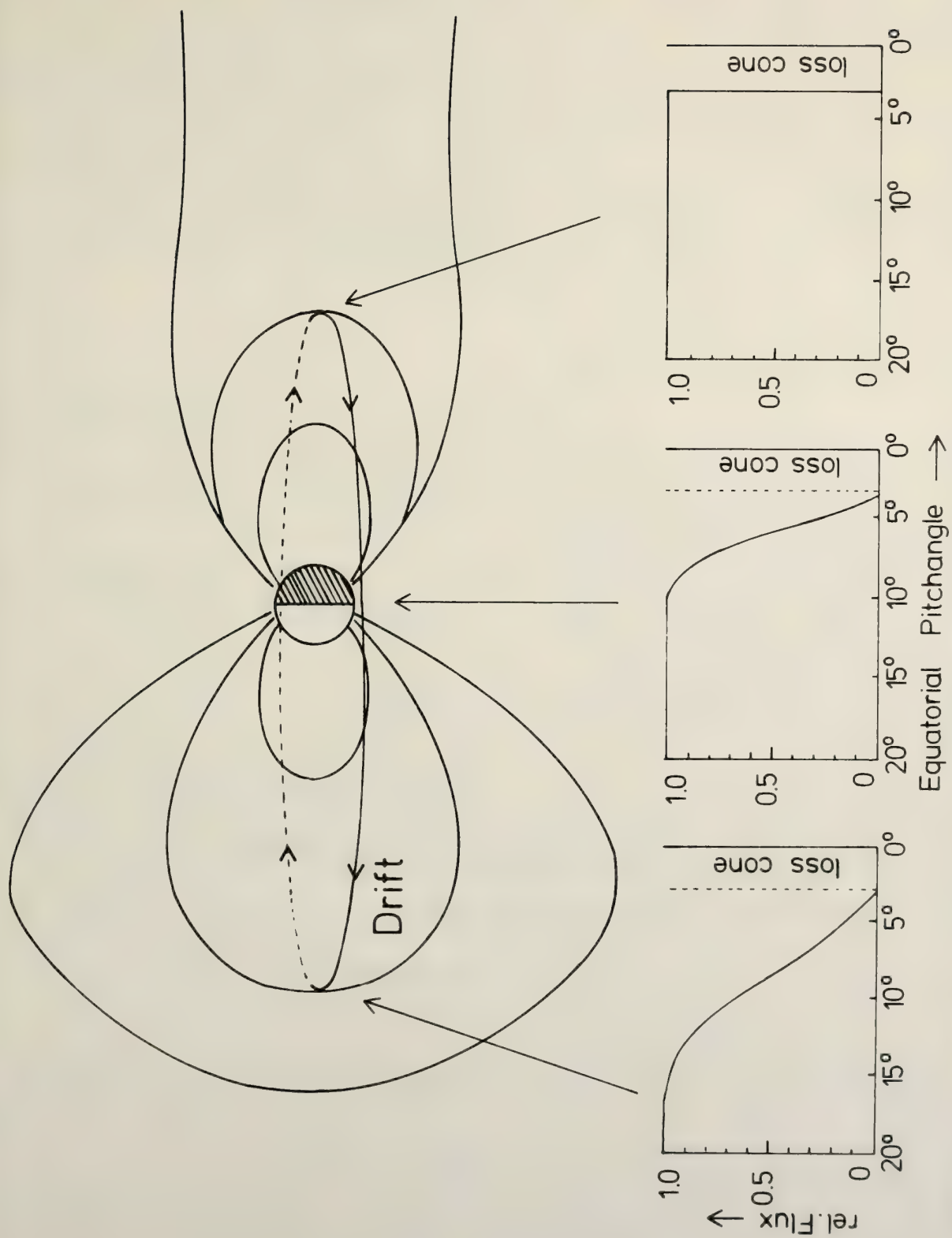


Figure 19. Schematic of the evolution of the equatorial pitch angle distribution under weak pitch angle scattering during the drift of a proton from dusk to dawn via noon.



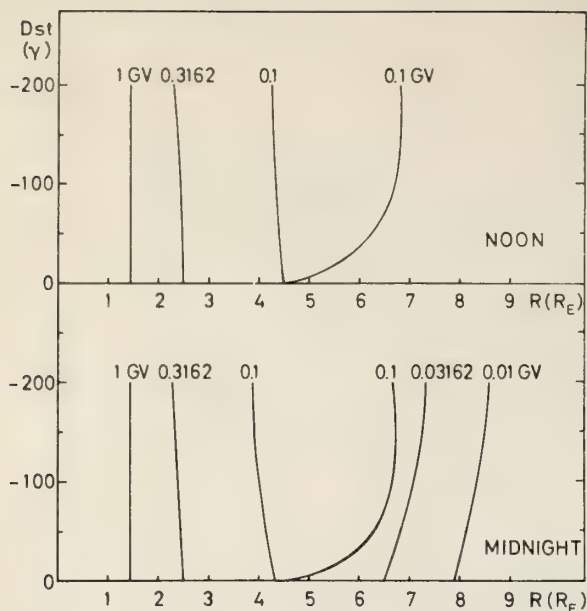


Figure 20. Computed iso-rigidity contours of breakdown of the guiding center approximation in a magnetosphere distorted by the solar wind and by the presence of a symmetric ring current. Contours are drawn in the equatorial plane and represent the radial distance from the geocenter as a function of Dst, both for noon and midnight meridional cross-sections (Morfill, 1974).

iso-rigidity contours of breakdown of the guiding center approximation in a magnetospheric field model including a symmetric ring current as a function of  $D_{st}$  (Morfill, 1974). Addition of a ring current does not affect the position of the radial iso-rigidity contours too much. However, due to the ring current inflation of the magnetospheric field lines originating at the same point on the earth reach further out and thus breakdown of the guiding center approximation occurs at a lower invariant latitude.

#### REFERENCES

- Blake, J. B., and L. M. Friesen (1977): 15th Int. Cosmic Ray Conf., 4:211.
- Durney, A. C., G. E. Morfill and J. J. Quenby (1972): J. Geophys. Res., 77:3345.
- Evans, L. C. (1972): Ph.D. Thesis, Calif. Inst. of Technology, Pasadena.
- Fennell, J. (1973): J. Geophys. Res., 78:1036.
- Gall, R., D. F. Smart and M. A. Shea (1971): Planet. Space Sci., 19:1419.
- Haskell, G. P., and R. J. Hynds (1972): In: Earth's Magnetospheric Processes, B. McCormac (ed.), D. Reidel Pub. Co., Dordrecht-Holland, p. 81.
- Hynds, R. J., G. Morfill and R. Rampling (1974): J. Geophys. Res., 79:1332.
- Montgomery, M. D., and S. Singer (1969): J. Geophys. Res., 74:2869.
- Morfill, G. (1973): J. Geophys. Res., 78:588.
- Morfill, G. (1974): In: Correlated Interplanetary and Magnetospheric Observations, D. E. Page (ed.), D. Reidel Pub. Co., Holland.
- Morfill, G., and J. J. Quenby (1971): Planet. Space Sci., 19:1541.
- Palmer, I. D., J. D. Jackson and E. W. Hones (1979): J. Geophys. Res., 84:109.
- Paulikas, G. A. (1974): Rev. Geophys. Space Phys., 12:117.

- Scholer, M. (1972): J. Geophys. Res., 77:2762.
- Scholer, M. (1975a): Space Sci. Rev., 17:3.
- Scholer, M. (1975b): Planet. Space Sci., 23:1445.
- Scholer, M., and G. Morfill (1975): J. Geophys. Res., 81:1737.
- Shea, M. A., D. F. Smart and K. G. McCracken (1965): AFCRL Tech. Rpt. 65-705.
- Williams, D. J., and F. T. Heuring (1973): J. Geophys. Res., 78:37.
- Williams, D. J., and G. D. Mead (1965): J. Geophys. Res., 70:3017.

## A REAL-TIME ISEE DATA SYSTEM

Bruce T. Tsurutani  
Jet Propulsion Laboratory  
California Institute of Technology  
Pasadena, California 91103

Daniel N. Baker  
University of California  
Los Alamos Scientific Laboratory  
Los Alamos, New Mexico 87545

Prediction of geomagnetic substorms and storms would be of great scientific and commercial interest. A real-time ISEE data system directed toward this purpose is discussed in detail. Such a system may allow up to 60+ minutes advance warning of magnetospheric substorms and up to 30 minute warnings of geomagnetic storms (and other disturbances) induced by high speed streams and solar flares. The proposed system utilizes existing capabilities of several agencies (NASA, NOAA, USAF), thereby minimizing costs. This same concept may be applicable to data from other spacecraft, and other NASA centers, thus allowing each individual experimenter to receive "quick-look" data in real time at his or her base institution.

### INTRODUCTION

It would be of great potential benefit for scientific, commercial and national defense purposes to be able to predict the onset of a magnetospheric substorm or a geomagnetic storm. The Solar Terrestrial Predictions Workshop has shown that if the solar wind plasma and fields which are about to impinge upon the Earth's magnetosphere are known, it may be possible to predict the intensity, duration, and other features of such geomagnetic activity.

At the present time there exists a very good opportunity to implement a real-time interplanetary (IP) monitoring system to make such geomagnetic predictions. The heart of this system is the third International Sun-Earth Explorer spacecraft (ISEE-3) which is located some  $240 R_E$  ( $1 R_E = 6375 \text{ km}$ ) upstream from the earth in a halo orbit (Figure 1) about the sun-earth libration point,  $L_1$ .<sup>(1)</sup> The ISEE-3 spacecraft has 12 onboard instruments which continuously measure the interplanetary medium.<sup>(2)</sup> In particular, ISEE-3 carries a sensitive magnetometer (provided by the Jet Propulsion Laboratory)<sup>(3)</sup> and a solar wind plasma instrument (provided by the Los Alamos Scientific Laboratory).<sup>(4)</sup> These instruments can, respectively, provide interplanetary magnetic field (IMF) and solar wind data that are indicative of interplanetary conditions well upstream of the earth.

For a quiet solar wind, with typical bulk flow speeds of the order of 300-400 km/sec, the interplanetary features measured at ISEE-3 will reach the earth's vicinity in about one hour's time. Any associated growth phase of



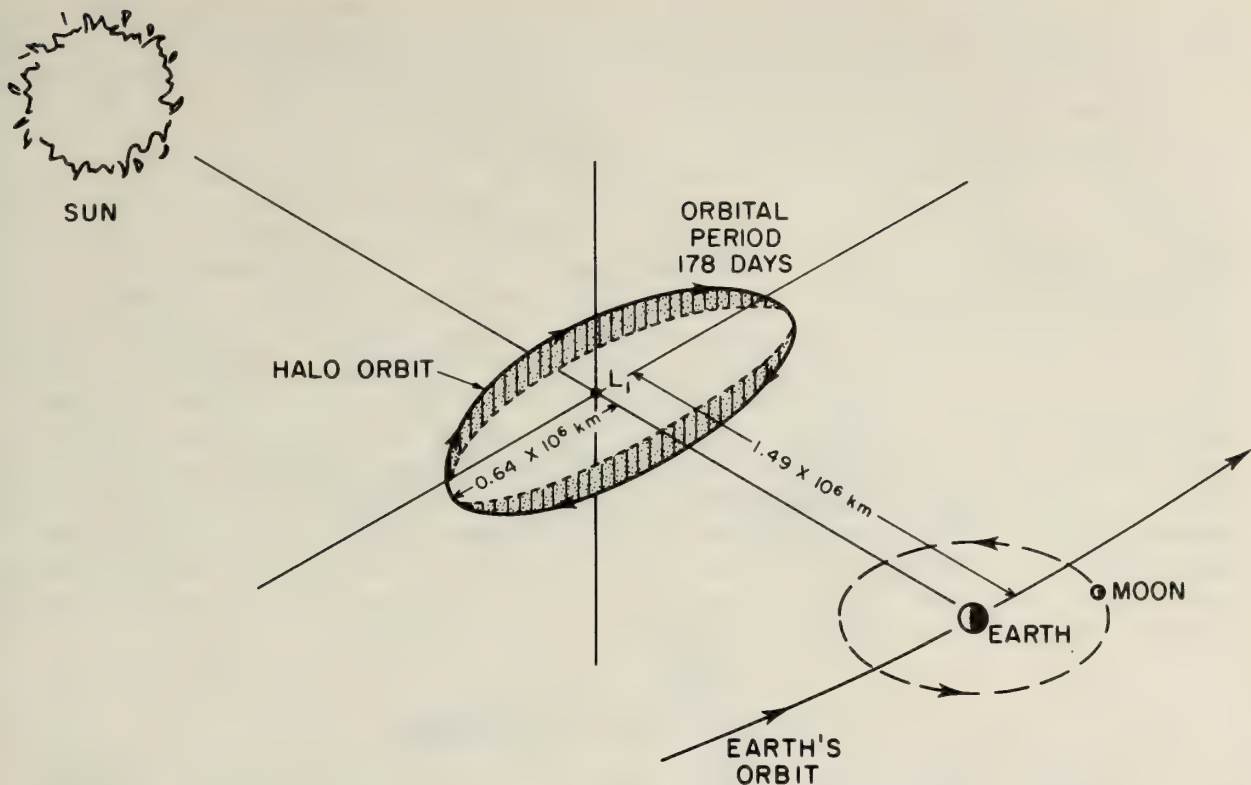


Figure 1. The International Sun Earth Explorer-3 orbit about the sun-earth libration point,  $L_1$ .

substorms<sup>(5)</sup> will give a further predictive lead time for such cases. However, velocities of 600 km/sec, or even greater, occur in high speed corotating streams<sup>(6)</sup> associated with coronal holes, implying a lead time of only ~30 minutes for recurrent events. Geomagnetic storm-time warnings might be even shorter owing to the higher solar wind streaming velocities which occur in association with the solar flares that cause storms. A real-time data system employing data from a satellite closer to the sun used in conjunction with either magnetohydrodynamic numerical modeling<sup>(7)</sup> or field-line tracing<sup>(8)</sup> may provide additional warning times in such cases. Furthermore, if the ISEE-3 solar x-ray instrument is also monitored<sup>(9)</sup>, longer times would be available for events in these latter categories.

Our discussion of predictions in this paper will center around substorm predictive capabilities. Various workers have shown that a host of parameters in the solar wind are (or may be) important indicators of magnetospheric substorm response. Among these interplanetary parameters are:  $V_{sw}$ , the interplanetary magnetic field z-component ( $B_z$ ), and combinations of  $V_{sw}$  and the interplanetary magnetic field such as  $V_{sw}B_z$ ,  $V_{sw}B_y$ , etc.<sup>(10)</sup> At present, no one has had available a continuous, long-term, real-time set of interplanetary data in order to truly evaluate a priori the predictive character of these parameters. All analyses to date have been done with archival data evaluated a posteriori. Now, with the advent of ISEE-3 and a strong interest in solar-terrestrial predictions, a true magnetospheric predictive capability can be developed.

Four agencies, viz., NASA, NOAA, the U.S. Air Force, and NSF, have expressed interest in using (and possibly funding) real-time ISEE data to make predictions. It is expected that predictions made under such a system would also be made available to the general scientific community.

We envisage that the International Magnetospheric Study (IMS) and related scientific activities could benefit from a real-time substorm prediction capability. For example, rocket and balloon launches could be keyed to the output of such a predictive program. Moreover, satellite experiment mode changes, spacecraft pointing changes, etc., could be implemented so as to optimize scientific payoff. Air Force workers could make use of ISEE data to analyze the propagation of high-energy solar flare particles and could use the substorm predictive capabilities to prepare for ionospheric disturbances which might affect high frequency communications or various radar systems. Finally, commercial interest in magnetospheric phenomena should not be overlooked. Exploratory activities of oil companies using magnetometer techniques are adversely affected by geomagnetic storms and substorms. Also, civilian communication, both ground-based and via satellite, can be affected by magnetospheric disturbances. Because of possible exposure of passengers and crews, even airlines flying over the polar regions of the earth have an interest in solar and magnetospheric radiation conditions.

#### POSSIBLE REAL-TIME SYSTEMS

In Figure 2 we show a different sketch of the relationship between the sun, the earth, and the ISEE-3 spacecraft. The halo orbit about  $L_1$  is roughly elliptical with a semimajor axis of  $0.64 \times 10^6$  km. The orbit is such that

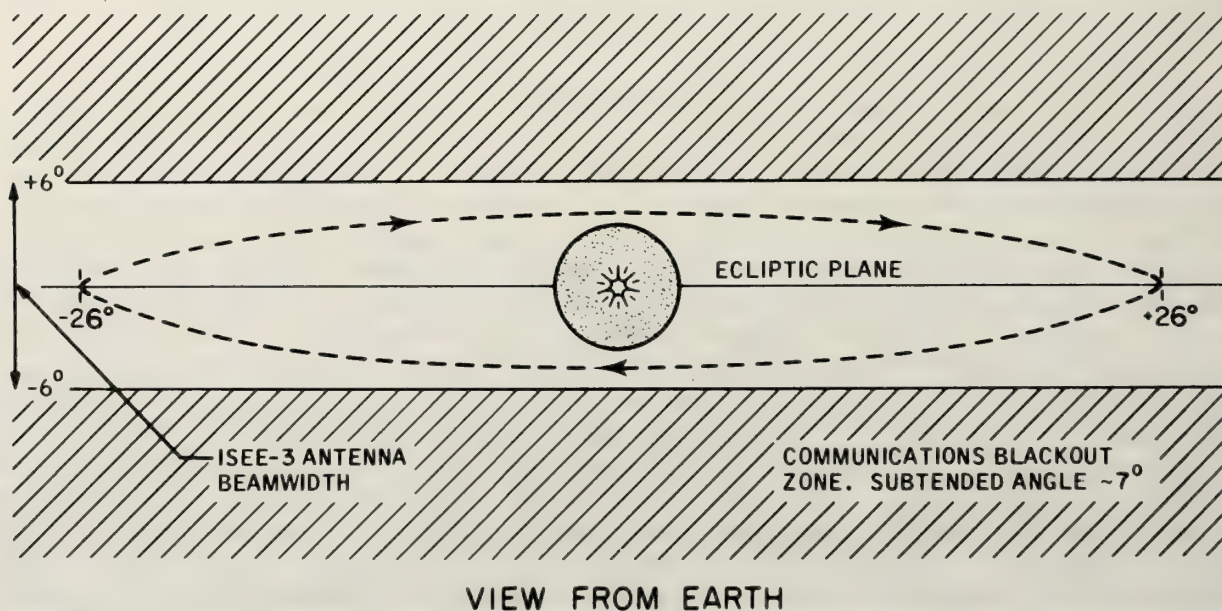


Figure 2. The ISEE-3 orbit as viewed from the Earth. The orbit avoids the region of strong solar radio interference, within about  $3.5^\circ$  of the sun, shown in the center of the figure. The ISEE spin axis is oriented perpendicular to the ecliptic plane. The full width transmission beam width is  $12^\circ$ , thus the ISEE orbit is constrained to within  $6^\circ$  of the ecliptic plane. (1)



when viewed from the earth, the spacecraft never moves across the face of the sun. If it did so, radio communication with ISEE-3 would be lost due to solar interference.<sup>(1)</sup> The ISEE orbital period is 178 days and at its maximum distance from the earth, the one-way light time is  $\sim 5.5$  seconds.

We remark here that it is possible that the substorm predictive capabilities of ISEE-3 may be degraded somewhat by the highly elliptical orbit. The scale length of both the field and plasma in the y- and z- direction must be calculated (perhaps by a correlative study of ISEE-1 and -3 data) to determine the correctness of the assumption that the characteristics of the solar wind as detected by ISEE-3 are the same as those impinging upon the magnetosphere. If the assumption is found to be incorrect, an assessment of the degradation of the predictability of substorms must be made.

Figure 3 illustrates additional elements in the ISEE-3 system as they exist or are presently conceived. ISEE data are telemetered to ground receiving stations on an S-band communications link. Computers at the ground stations send the received data to the Goddard Space Flight Center (GSFC) via the NASA communications (NASCOM) system in real time. The data received at a

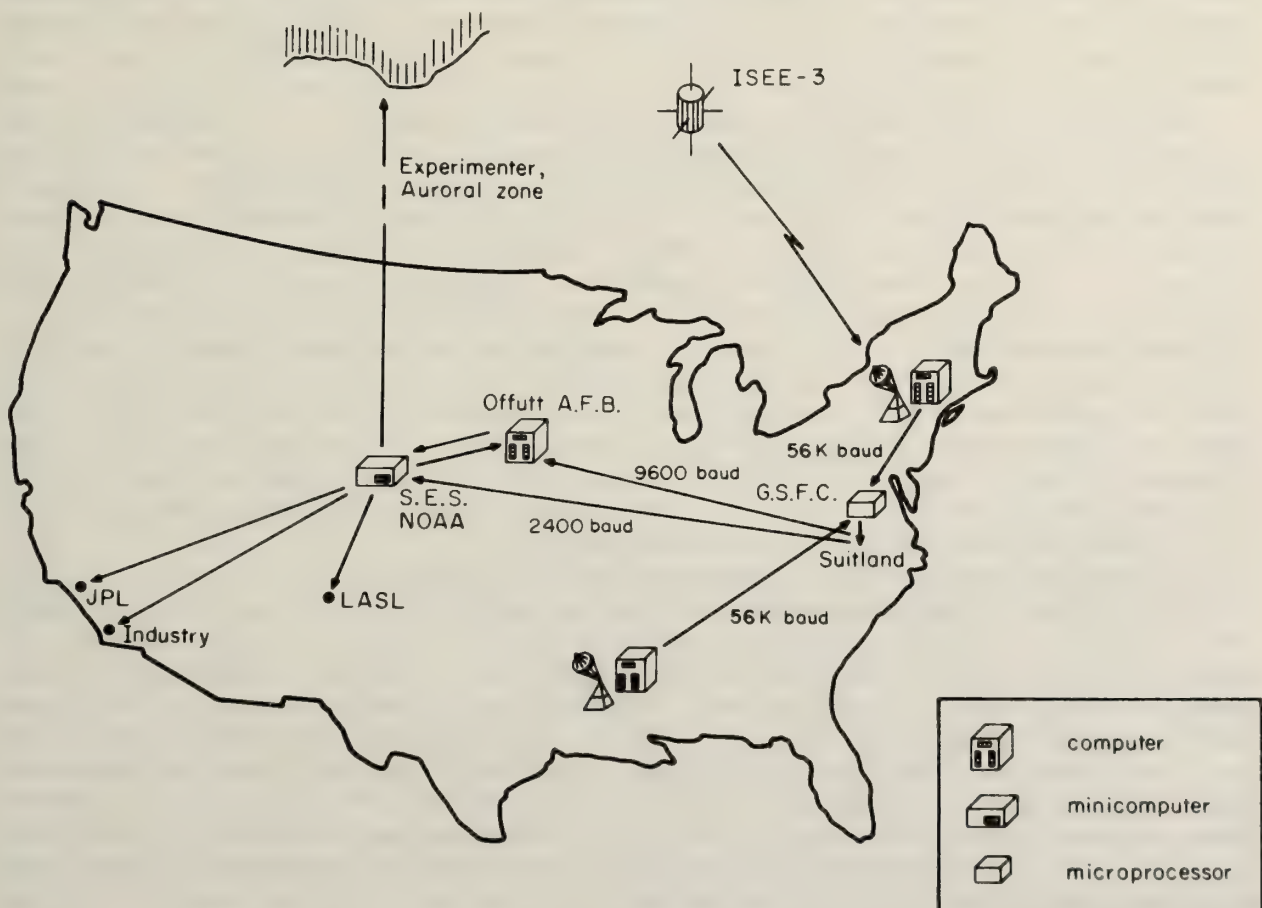


Figure 3. The ISEE real-time data link as it is presently envisioned. The encoded data are deconvolved at the ground station, then sent to G.S.F.C. via high data rate lines. The data are intercepted as they come into G.S.F.C. and are routed through Suitland, Maryland to the Space Environment Services Center, N.O.A.A. where the data reduction and index calculations are performed. Index dissemination will take place from the S.E.S.C. at Boulder, Colorado.



given ground station from the ISEE spacecraft are also recorded on tape. However, these tapes are never sent to GSFC. If NASCOM data transmission was not of adequate quality on the first transmission, the recorded data from the ground station are simply played back and sent to GSFC again by way of the NASCOM link.

NASCOM handles data from many satellite systems beside ISEE-3. This system also handles ISEE-1 and -2 data, for example. Hence, the NASCOM link is a 56 kbps (kbps =  $10^3$  bits per second) line capable of carrying many transmissions to Goddard from a variety of spacecraft being tracked by the ground station.

At Goddard, the data being carried by NASCOM can go to one of two locations. The first is the Multi-Satellite Operations Control Center (MSOCC) where the spacecraft location and operational status are monitored and commands are originated. The second location at Goddard is the Information Processing Division (IPD). The IPD has the responsibility for receiving the scientific data, reformatting it, and eventually sending data tapes to experimenters. A basic part of the IPD is a large mass storage capability called TELOPS. TELOPS has the potential to store all data from all NASCOM-linked spacecraft for 6 months. It also has an extensive editing capability which facilitates the production of final, "cleaned-up" data tapes for experimenters.

Because of the vast quantities of data entering the TELOPS system, the necessity of editing the final ISEE data output, and the precise ranging and timing requirements of the gamma-ray burst experiment on ISEE-3, it takes a substantial period of time (i.e., several weeks) for experimenter data tapes to finally be generated from TELOPS. Thus, for purposes of a real-time system, it is most practical to intercept the ISEE-3 data as it comes into GSFC on the NASCOM link and before it goes into TELOPS (T. Von Rosenvinge, personal communication). Hence, it is presently proposed that a suitable microprocessor-based computer be interposed into the system at Goddard and this microprocessor will be designed to recognize ISEE data, strip out and read the solar wind plasma and magnetometer data, and then transmit these relevant 700-800 bps to a data reduction site. The basic cost for such a microprocessor system to read and divert ISEE data in the NASCOM stream is ~\$50K, with NASA providing the necessary funds.

Having, by means of the GSFC microprocessor system, isolated and diverted the ISEE-3 data stream necessary for magnetospheric predictions, there comes the problem of further processing the raw data to make it useful in physical terms. In the case of the LASL solar wind plasma data, the basic instrumental counting rates must be converted into moments of the distribution function such as density, velocity, etc. In the case of the JPL magnetometer, the data must be "despun" to get interplanetary magnetic fields in terms of the usual solar magnetospheric cartesian components. Even for relatively rough cuts at these reduced parameters, fairly complicated algorithms are required and substantial computer processing capability is necessary. We presently see three possible locations for such processing: (1) at GSFC; (2) at Offutt A.F.B., Nebraska; and (3) at NOAA in Boulder, Colorado.

The GSFC-processing option looks very desirable in many ways. The diverted ISEE data stream is immediately available at Goddard and this facility has a long history of handling such data, overseeing preliminary data reduction, and protecting the interests of the principal investigators (PIs). However, GSFC at present does not have a minicomputer in place to do the processing being discussed and it does not have the available manpower to take



care of these tasks around the clock. Thus, with the Goddard-processing option, we are talking about the requirement for new hardware (a minicomputer: ~ \$20K) and new personnel. Given these requirements, the GSFC option looks expensive and less desirable.

The second option involves data processing at the U.S. Air Force Air Weather Service (AWS) facility at Global Weather Central, Offutt A.F.B. in Omaha, Nebraska. This option obviously has an immediate drawback, namely, the distance between GSFC and Omaha. The 700-800 bps stripped out by the micro-processor at Goddard must be shipped continually to AWS for this to be a workable real-time approach. Fortunately, and as illustrated by Figure 3, there presently exists a high-speed (9600 baud) line most of the way between Goddard and Offutt. This line extends between Suitland (the NOAA National Environmental Satellite Center) in Maryland and Offutt Field. Thus with a relatively modest cost (~ \$0.30 per mile) of establishing a high-speed line from GSFC to Suitland, the Offutt-processing option would be possible.

The Air Weather Service has had considerable experience in handling, reducing, and utilizing real-time data from satellite systems, as well as protecting the interests of investigators. This experience includes many years of receiving LASL solar wind plasma data from the Vela series of spacecraft. The AWS also has in place computers which are manned at all times. However, the AWS computers are presently operating near their capacity and AWS is only marginally staffed to accomplish its present space environmental monitoring services. Current Air Force funding precludes adding more AWS hardware capability now or in the near future and additional staffing must await a requirement directive from the Air Force. Thus with the AWS-option, as with the GSFC-option, hardware and personnel costs detract from the present use of this avenue for a real-time ISEE-3 prediction system.

This brings us to the third option, which involves sending the raw ISEE-3 data to the Space Environment Services Center (SESC), operated by NOAA in Boulder, Colorado. This is a desirable approach since the SESC group has had long experience in the collection and dissemination of data from space systems. This option again requires the emplacement of a high-speed line from GSFC to Suitland, as was required for option (2) above. There presently exists a 2400 baud line between Suitland and SESC, as is shown in Figure 3. Thus, transmission of ISEE data to Boulder presents no major obstacle and, furthermore, the SESC personnel have expressed a desire and capability, as described below, to handle ISEE data in a fashion directed toward real-time alert usage.<sup>(11)</sup>

The essential features of the NOAA Space Environment Laboratory data system (SELDADS) are illustrated in Figure 4. The system is composed of two Data General NOVA and one Data General ECLIPSE minicomputers with 58K Megabyte storage capability for each NOVA and smaller storage on the ECLIPSE. At present, data from x-ray, particle, and magnetometer sensors on two GEOS satellites, the TIROS-N spacecraft, from ground-based IMS magnetometers, and from various teletype networks are already being collected by this system. The ISEE-3 data would be an additional load on the SELDADS system and the computer processing and programming resources available to SELDADS would be the principal limitation to handling ISEE data. As illustrated in Figure 4, ISEE data would be received by the NOVA minicomputers and the data would be recorded on a disk storage device. Subsequent processing of the data (as described below) would then provide reduced interplanetary data (or indices) to public users and SESC forecasters via the ECLIPSE computer.

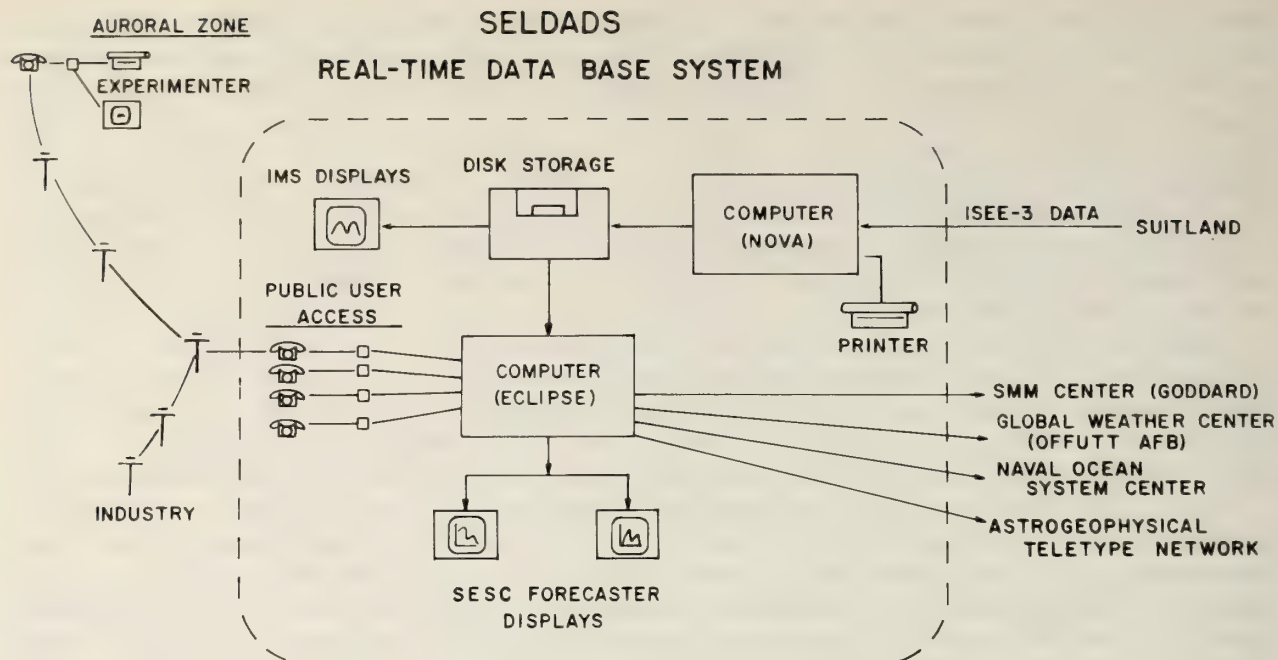


Figure 4. The NOAA Space Environment Laboratory Data Acquisition and Display System (SELDADS) which will be used to reduce the data, form the substorm and storm predictive indices, and disseminate this information.

Because SELDADS is presently taxed in terms of both hardware capability and programmer support, SESC personnel have suggested a phased approach to implementing the ISEE real-time system:

Phase I Data would be handled by minimal programs that would extract a data point every one minute or five minutes (depending on complexity of the next step), compute the 3-dimensional magnetic field parameters and bulk plasma parameters and output these on a printer. These would be monitored visually by a forecaster for configurations of interest that would initiate an alert and that would be relayed verbally to GWC (Global Weather Central). Phase I could be implemented as soon as the microprocessor system is installed at Goddard, the Goddard-Suitland link is provided, and some programming is done at Boulder.

Phase II As some programming time and space become available at Boulder, the data from Phase I would be added to the data stream flowing to GWC on the 2400 baud data line. At the same time, it would be put in a file in the SELDADS for driving some simple CRT displays in the SESC. This could take from 6 months to a year after Phase I is completed.

Phase III During this phase an implementation of more sophisticated displays and algorithms that produce indices yet to be determined could occur. Phase III would come at some undefined time following Phase II.



Table 1. The ISEE-3 real-time data flow as it is presently envisioned. See Figure 3.

FLOW DIAGRAM

<u>Facility</u>	<u>Location</u>	<u>Hardware</u>	<u>Data Processing</u>	<u>Transmission Link</u>
ISEE-3	Earth-Sun Liberation Point, L1. Halo Orbit 0.01 AU from Earth	VHM Magnetometer  Plasma Analyzer		S-band
Ground Station		Computer	Deconvolve Data	56 K Baud Line
NASCOM	Goddard S. F. C. Greenbelt, MD	Heetderks/Mozer Black Box (microprocessor)	Strip off ISEE-3 Magnetometer, Plasma Data	Dedicated Phone Line
N.O.A.A.	Suitland, MD		None	Dedicated Phone Line
Space Environ- ment Services Center, N.O.A.A.	Boulder, CO	Minicomputer Data Displays	Despin Data to Inertial Coor- dinates. Remove Spacecraft Offsets. Calculate Velocity. Calculate Substorm. Prediction Indices	Two-Wire Telephone (to Experimenter, Industry) Computer-Computer Link (to Offutt)
Experimenter Industry		Modem		
Offutt A.F.B.	Omaha, NE			

Figure 3 and Table 1 illustrate the essential features of the flow of ISEE-3 data in the proposed real-time system. Beginning at the spacecraft, the data from all instruments are telemetered to a particular ground station. After initial deconvolution, the data are sent to Goddard SFC via the NASCOM 56K baud line. At Goddard a microprocessor system strips off the ISEE-3 solar wind and IMF data and these data are then sent via Suitland to the SESC facility illustrated in Figure 4 above. Finally, after processing at Boulder, substorm alerts and predictive indices (yet to be chosen) are sent out and provided to IMS experimenters, to industrial users, and to the military.

CONCLUSIONS

We conclude from our own investigations, and from the generous input of many other knowledgeable persons, that it is readily within present technical capabilities to implement a real-time prediction system using upstream interplanetary data input. As outlined above, a crude system (Phase I) could be established within a matter of a few months. A more sophisticated system involving the generation of CRT displays and well-conceived predictive parameters will probably take on the order of one to two years to implement. As presently envisaged, the real-time ISEE-3 system being proposed here can be implemented at a modest cost of perhaps \$50-100K.

As described in the Introduction, many benefits can accrue from the presently proposed real-time system. Scientific, commercial, and military activities will all be enhanced, possibly in many unforeseen ways, by knowing more about the near-earth space environment, and knowing this in advance. Thus, we may expect to benefit scientifically and socially from the modest investment required in order to implement a real-time alert system.

We have stressed an ISEE-3 real-time substorm prediction system in our present discussions, but obvious extensions of these ideas spring to mind. First of all, we can envision providing all ISEE-3 experimenters with their data on a real-time basis. As illustrated by Figure 5, if one once again had the proper microprocessor system located at Goddard to tap into the NASCOM system, one could easily send the ISEE-3 data stream of interest to JPL, LASL, or wherever needed. A phone link, appropriate modems, and a minicomputer to process the data would together provide a much improved "quick-look" data system. Something of this sort could be done at present even if a substorm prediction system were not implemented.

Obviously, the above quick-look system need not be restricted to ISEE-3. The same basic approach could be used for ISEE-1 and -2, or any other NASA spacecraft data which are being sent via the NASCOM link. The same approach

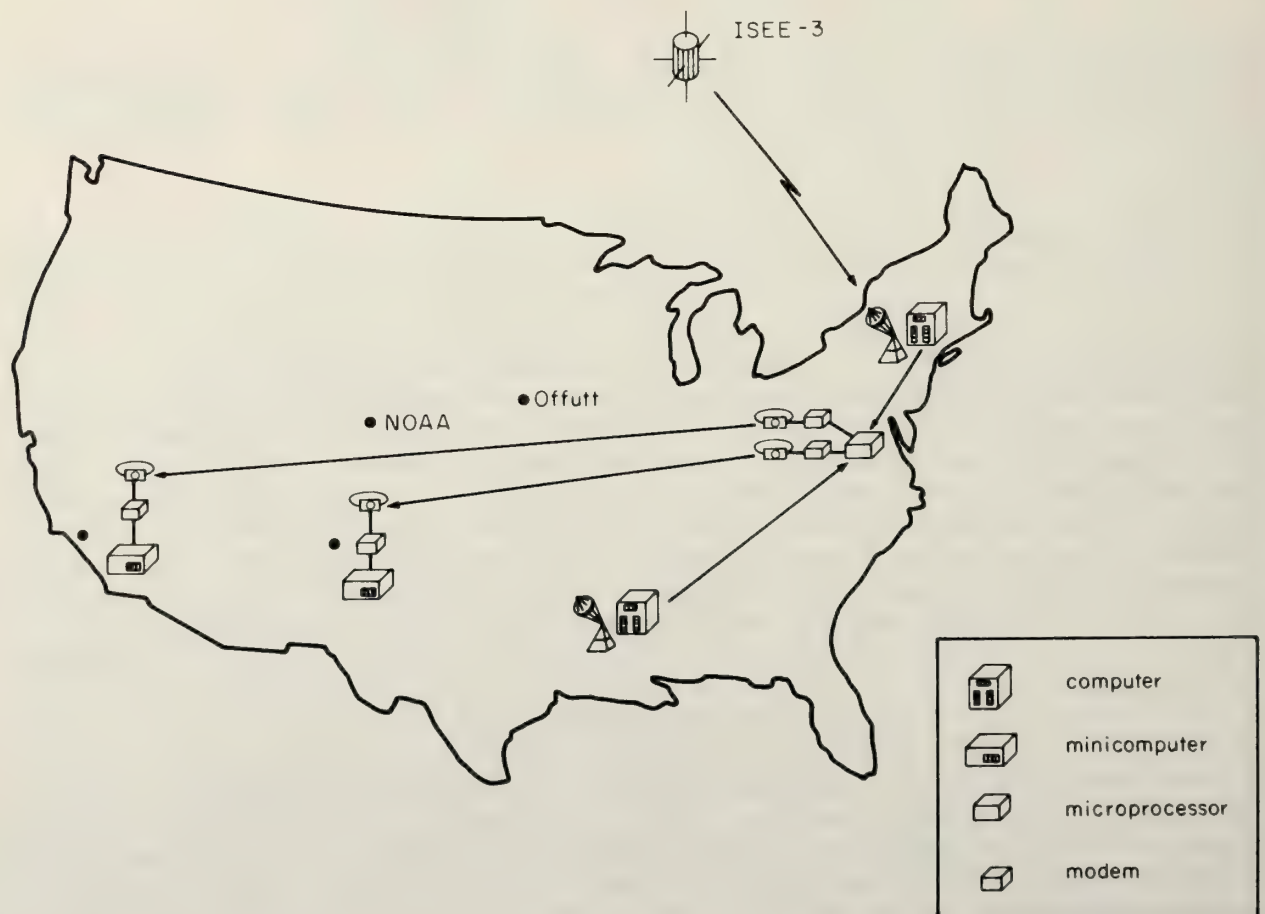


Figure 5. An ISEE real-time "quick look" data system. This same approach could be used to send "quick look" data from all NASA satellites to individual experimenters.

KEY PERSONNEL

National Aeronautics and Space Administration Headquarters

H. Glaser  
E. Schmerling  
M. Wiskerchen

Air Force

Major V. Patterson  
Lt. Col. L. Snyder  
Lt. Col. R. Thompson

National Science Foundation/International Magnetospheric Survey

R. Manka

National Oceanic and Atmospheric Administration

G. Heckman  
C. Hornback  
D. Williams

Goddard Space Flight Center

R. Farquhar  
D. Muhonen  
T. Von Rosenvinge

ISEE-3 Experimenters

D. Baker  
S. Bame  
E. Smith  
B. Tsurutani

Solar Terrestrial Physics Working Group

S. Akasofu  
R. Donnelly  
G. Paulikas  
C. Russell

Figure 6. Some of the key people involved in constructing the ISEE-3 real-time data system.



could also easily be extended to other NASA centers because the difficult, initial data deconvolution is done by computers at the ground stations. Thus the ideas being explored here may represent the wave of the future as to how NASA and various satellite experimenters exchange information and preliminary data.

#### ACKNOWLEDGMENTS

The ideas and information presented in this paper are the result of the efforts of many people. Figure 6 shows a list of key personnel at various institutions who have and, hopefully, will continue to play instrumental roles in the real-time ISEE-3 system. We are particularly indebted to Sam Bame and Ed Smith who are, respectively, the principal investigators on the ISEE-3 plasma analyzer and Vector Helium magnetometer experiments. We also thank Tycho Von Rosenvinge, the ISEE-3 project scientist, for patient descriptions of the ISEE data system and we thank Mike Wiskerchen and Erwin Schmerling for pursuing the details of the data system, and most importantly for obtaining NASA funds to make this possible. We sincerely thank Gary Heckman and Don Williams of the SESC in Boulder for providing us with detailed information about the SELDADS system and we acknowledge many useful conversations with Dick Thompson and Vern Patterson of the Air Weather Service regarding the Air Force role in the utilization of ISEE data. Finally, we thank Dick Donnelly and the chairmen of the International Solar-Terrestrial Predictions working groups (Syun Akasofu, George Paulikas, and Chris Russell) for providing the forum in which to discuss the ideas presented in this paper.

This report represents one aspect of research carried out by the Jet Propulsion Laboratory for NASA under Contract NAS7-100. Work performed at the Los Alamos Scientific Laboratory was done under the auspices of the U.S. Department of Energy.

#### BIBLIOGRAPHY

1. Farquhar, R. W., D. P. Muhonen and D. L. Richardson, Mission Design for a Halo Orbiter of the Earth, J. of Spacecraft and Rockets, 14, 170, 1977.
  2. Ogilvie, K. W., A. Durney and T. Von Rosenvinge, Descriptions of Experimental Investigations and Instruments for the ISEE Spacecraft, IEEE Trans. Geosci. Electr., GE-16, 151, 1978.
  3. Frandsen, A. M. A., B. V. Connor, J. Van Amersfoort and E. J. Smith, The ISEE-C Vector Helium Magnetometer, IEEE Trans. Geosci. Electr., GE-16, 195, 1978.
  4. Bame, S. J., J. R. Asbridge, H. E. Felthausen, J. P. Glore, H. L. Hawk, and J. Chavez, ISEE-C Solar Wind Plasma Experiment, IEEE Trans. Geosci. Electr., GE-16, 160, 1978.
  5. Baker, D. N., P. R. Higbie, E. W. Hones, Jr. and R. D. Belian, High-Resolution Energetic Particle Measurements at  $6.6 R_E$ . 3. Low-Energy Electron Anisotropies and Short-Term Substorm Predictions, J. Geophys. Res., 83, 4862, 1978.
- Clauer, C. R. and R. L. McPherron, Mapping the Local Time-Universal time Development of Magnetospheric Substorms Using Mid-Latitude Magnetic Observations, J. Geophys. Res., 79, 2811, 1974.

6. Smith, E. J. and J. H. Wolfe, Observations of Interaction Regions and Corotating Shocks Between One and Five AU: Pioneers 10 and 11, Geophys. Res. Lett., 3, 137, 1976.  
Gosling, J. T., A. J. Hundhausen and S. J. Bame, Solar Wind Stream Evolution at Large Heliocentric Distances: Experimental Demonstration and Test of a Model, J. Geophys. Res., 81, 2111, 1976.
7. Dryer, M., Z. K. Smith, E. J. Smith, J. D. Mihalov, J. H. Wolfe, R. S. Steinolfson and S. T. Wu, Dynamic MHD Modeling of Solar Wind Corotating Stream Interaction Regions Observed by Pioneer 10 and 11, J. Geophys. Res., 83, 4347, 1978.  
Steinolfson, R. S., M. Dryer and Y. Nakagawa, Numerical MHD Simulation of Interplanetary Shock Pairs, J. Geophys. Res., 80, 1223, 1975.
8. Roelof, E. C., Coronal Structure and the Solar Wind, in Solar Wind Three, ed. by C. T. Russell, 98, Univ. of Calif. Press, Los Angeles, 1974.
9. Kane, S., Solar Flare Monitoring Using the ISEE-3 X-Ray Instrument, Proc. of Solar Terrestrial Predictions Workshop, 1979.
10. Tsurutani, B. T. and C. I. Meng, Interplanetary Magnetic-Field Variations and Substorm Activity, J. Geophys. Res., 77, 2964, 1972.  
Meng, C. I., B. Tsurutani, K. Kawasaki and S. I. Akasofu, J. Geophys. Res., 78, 617, 1973.  
Feynman, J., Substorms and the Interplanetary Magnetic Field, J. Geophys. Res., 81, 5551, 1976.
11. Heckman, G. R., A Summary of the Indices and Predictions of the Space Environment Services Center, Proceedings of the International Solar-Terrestrial Predictions Workshop, preprint 170, 1979.
12. Paulikas, G. A., Report of the USAF Scientific Advisory Board Ad Hoc Committee on Aeronomy, USAF Report, HQ USAF, Washington, D.C., 1977.

## VI. IONOSPHERIC PREDICTIONS

### A. MAGNETOSPHERE-IONOSPHERE INTERACTIONS

A WORKING GROUP REPORT prepared by R. R. Vondrak, Chairman, Y. T. Chiu, D. S. Evans, V. G. Patterson, G. J. Romick and K. Stasiewicz.

#### 1. INTRODUCTION

The ionosphere and magnetosphere are two elements of an interactive system. They are coupled by physical processes occurring in regions that are partly overlapping, although these regions are often considered as spatially separated. Neither the magnetosphere nor the ionosphere is a completely independent system and changes in one may affect the other.

Processes that couple the magnetosphere to the ionosphere can produce considerable effects on both. For example, depending upon level of disturbance, anywhere from  $10^{10}$  to  $10^{12}$  watts is deposited in the earth's upper atmosphere by particle precipitation and Joule heating. By comparison, solar radiation incident on the earth is about  $1.7 \times 10^{17}$  watts, with only about  $6 \times 10^{11}$  watts absorbed at altitudes above 90 km. However, this energy deposition has significant effect in that it maintains the daytime ionosphere. Energy deposition during an auroral sub-storm is a substantial transient on the solar illumination that serves as a regular driving function on the upper atmosphere. Such a transient energy input can produce localized and relatively short-lived perturbations in the ionosphere and upper atmosphere. The most important of the transient energy injections are traceable to processes in the earth's magnetosphere, although it is recognized that the ultimate driving function for these processes lies outside the magnetosphere in the solar wind.

The purpose of this report is to describe present understanding of magnetosphere-ionosphere interactions and to assess present and future predictive capabilities. Because several topics are also of interest to other working groups, we will primarily consider those ionospheric features that are directly coupled to the magnetosphere to a significant degree. Emphasis will be given to those phenomena that are of major interest to forecasters and users.

In the following sections of this report we first identify and describe the interactive components of the magnetosphere-ionosphere system. Next, we identify the elements that are of greatest relevance to users and forecasters. Monitoring and prediction techniques are described in the next section, and are followed by a discussion of goals and expectations for future progress. We conclude with specific recommendations for achieving progress in understanding ionosphere-magnetosphere interactions.

#### 2. MAGNETOSPHERE-IONOSPHERE INTERACTIONS

Interactions between the magnetosphere and ionosphere can be divided into three basic categories: the transfer of energy from the magnetosphere to the ionosphere, electrical coupling by electric fields and electric currents, and the exchange of plasma. In this section we will identify the ionospheric regions associated with these interactions, describe the physical processes that occur, and summarize some of the recent progress in our understanding.

Various aspects of magnetosphere-ionosphere interactions have been discussed in recent reviews that are useful for a more detailed description and complete listing of references (e.g., Akasofu, 1977; Bostrom, 1972; Furch, 1977; Chiu, 1979; Kamide, 1979; Wolf, 1975).



## 2.1 Energy Input

The injection of energy from the magnetosphere into the ionosphere takes two forms: the precipitation of energetic particles into the atmosphere, and the heating of the neutral atmosphere by dissipative ionospheric currents that are driven by magnetospheric dynamo processes. Together these two energy sources globally input typically  $10^{10}$  to  $10^{12}$  watts. This is smaller than the global solar radiation input of  $1.7 \times 10^{17}$  watts, but is comparable to the  $6 \times 10^{11}$  watts of solar radiation absorbed above 90 km. Because magnetospheric energy is almost totally deposited at altitudes above 90 km and at latitudes higher than  $55^\circ$  ( $< 20\%$  of the earth's area), it can profoundly affect the high-latitude upper atmosphere. Indeed, the energy deposited into the high-altitude arctic atmosphere by the magnetosphere often dominates that from the sun and so affect thermospheric dynamics as to alter the low-latitude thermosphere and ionosphere.

The energy input tends to be dissipated in two different forms. That carried by the energetic particles goes primarily into the production of ionization, although after ion recombination and radiation it eventually heats the neutral gas. On the other hand, that energy dissipated by the ionospheric currents is transferred directly to the neutral atmosphere and appears in the form of both heating and directed motion of the air mass. In terms of relative magnitudes, about 30% of the average energy input is by charged particle precipitation and the remainder is by dissipative current flow.

### 2.1.1 High-Latitude Particle Precipitation

At high latitudes the precipitation of energetic particles typically occurs in an oval region about the geomagnetic poles at geomagnetic latitudes between  $60^\circ$  and  $80^\circ$ . These particles (electrons and protons) originate in the magnetospheric plasma sheet and are energized to several kilovolts by processes still not well understood.

The pattern of auroral particle precipitation is shown in Figure 1. In the premidnight sector bright arcs occur poleward of a broad band of diffuse aurora. In the morning sector the precipitation is more structured. The morphology shown in Figure 1 is representative of conditions during an auroral substorm. A substorm is a sequence of events that usually begins with the sudden brightening of a quiet arc in the midnight sector. Subsequently, the arc moves poleward rapidly, forming a bulge in the midnight sector that then expands in all directions. In the evening side this expansion is characterized by a westward traveling surge, and on the morning side the arcs disintegrate into patches. After this expansion the aurora gradually recovers to its pre-substorm quiescent state. The entire duration of an auroral substorm is typically 1 to 2 hours.

Because the kilovolt auroral electrons are the primary source of high-latitude E-region ionization, the pattern of ionospheric electron density closely resembles that shown in Figure 1.

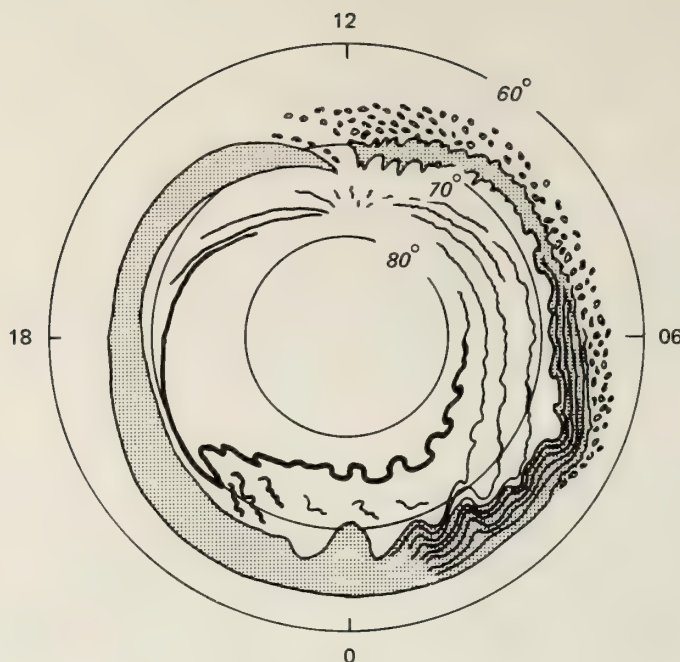


FIGURE 1 SCHEMATIC DIAGRAM SHOWING THE MAIN CHARACTERISTICS OF AURORAS DURING AN AURORAL SUBSTORM IN COORDINATES OF DIPOLE LATITUDE AND MAGNETIC LOCAL TIME. Discrete arcs are indicated by lines and the diffuse auroral regions are shaded (after Akasofu, 1976).

In addition to the kilovolt particles that produce the visual aurora, higher energy particles deposit energy at high-latitudes during REP (relativistic electron precipitation) and PCA (polar cap absorption) events. These relativistic particles primarily affect the D-region.

### 2.1.2 High-Latitude Currents

Thermospheric heating at high latitudes is closely associated with the system of horizontal currents (auroral electrojet). The auroral electrojet system was first detected by ground-based magnetometers. It is characterized by an eastward current before midnight and a westward current after midnight that intensifies during an auroral substorm (see Figure 2). The auroral electrojet system is approximately colocated with the regions of auroral precipitation, although there is evidence that the currents may be reduced within auroral arcs. The total current in the auroral electrojet is about  $10^6$  A and is driven by electrical coupling between the high-latitude ionosphere and the magnetosphere.

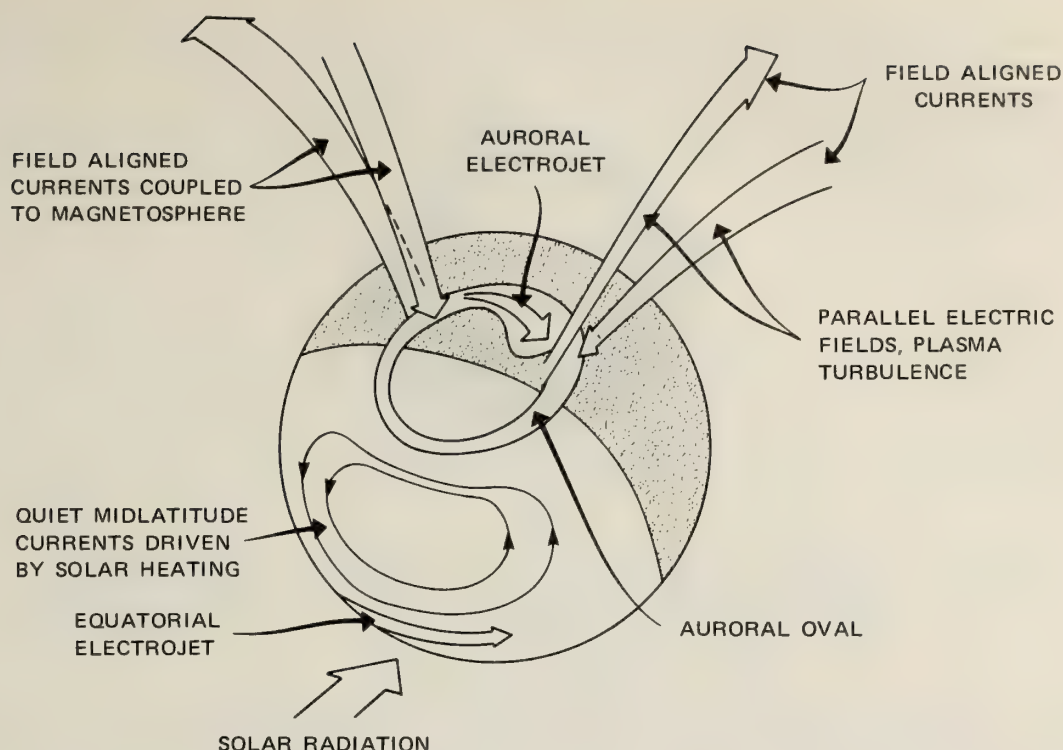


FIGURE 2 THE CONFIGURATION OF THE MAJOR CURRENT SYSTEMS IN THE EARTH'S IONOSPHERE. At high latitudes the auroral electrojet is coupled to the magnetosphere by field-aligned currents. The mid-latitude and equatorial currents are primarily driven by solar radiation heating affected by changes in the magnetospheric electric field (after Space Science Board, 1978).

## 2.2 Electrical Coupling

The magnetosphere and the ionosphere are electrically coupled by means of the high-latitude electric field and current systems.

### 2.2.1 Convection Electric Field

In the magnetosphere, plasma is driven towards the earth from the magnetotail by a large-scale electric field directed from dawn to dusk. This electric field is called the convection electric field and is due to the motion of the solar wind and the interplanetary magnetic field past the earth. The convection electric field is mapped along the geomagnetic-field lines into the high-latitude ionosphere producing a general anti-sunward convection pattern (Figure 3). The ionospheric plasma moves perpendicular to the ionospheric electric field, so the flow lines can be regarded as electric equipotentials.



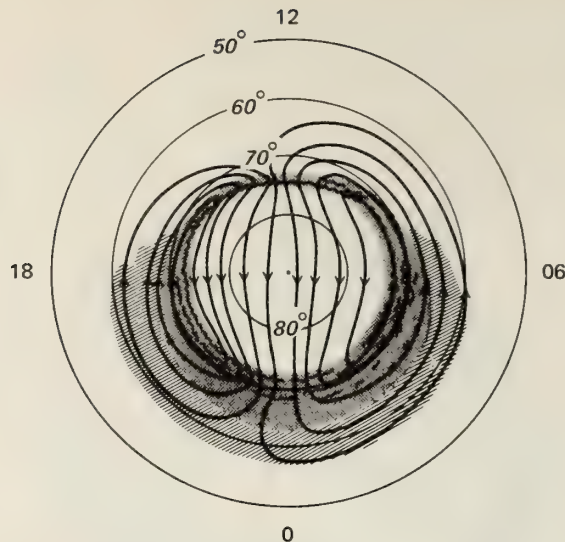


FIGURE 3 SCHEMATIC ILLUSTRATION OF THE HIGH-LATITUDE PLASMA CONVECTION TRAJECTORIES (heavy solid lines), DISCRETE AURORA (wavy lines), DIFFUSE AURORA (dotted area), AND PLASMA TROUGH (hatched area), (after Schunk et al., 1976).

The two-celled convection pattern in the ionosphere is characterized by a general dawn-to-dusk electric field in the polar cap, a northward electric field in the pre-midnight auroral zone, and a southward electric field in the morning auroral zone. Not well understood are the electric field patterns in the dayside auroral region near noon (the plasma cleft), the region near local midnight where the field rotates from north to south (the Harang Discontinuity), and the small-scale variation of the electric field near auroral arcs.

The convection pattern of ionospheric plasma is a primary determinant of the morphology of the high-latitude F-region. In winter, transport of ionization from the sunlit hemisphere maintains the polar ionosphere. The mid-latitude trough south of the auroral oval is thought to result from the recombination of ionization along drift paths that stay in the nighttime region for long periods of time.

Another source of electric fields at all latitudes is the motion of the neutral atmosphere. Winds that transport ionization across magnetic field lines serve as a dynamo producing an electric field that at high latitudes may couple to the magnetospheric plasma.

The convection electric field may also influence the low-latitude ionosphere. Under certain conditions the convection electric field may penetrate to the equator where it can significantly alter the low-latitude electric field and current.

### 2.2.2 Field-Aligned Currents

The high-latitude ionosphere and the magnetosphere are electrically coupled by means of field-aligned currents, (i.e., a flow of electrical charge along the geomagnetic field lines). These sheets of current originate in the geomagnetic tail and plasma sheet and extend both in latitude and local time along the auroral oval (see Figure 2).

The current density of the large-scale field-aligned current system is typically  $0.1$  to  $10 \mu\text{A}/\text{m}^2$ , with a total current flow into (and out of) the ionosphere of about  $2 \times 10^6 \text{A}$  during quiet times and  $5 \times 10^6 \text{A}$  during disturbed times. The general pattern of the field-aligned current system remains intact during increasing geomagnetic activity, but it shifts to somewhat lower latitudes.

### 2.3 Plasma Transport

The regions of the ionosphere in which plasma flows into or out of the magnetosphere are identified in Figure 4. The flow of energetic plasma into the high-latitude ionosphere is well known. However, lower-energy thermal plasma flows out of the ionosphere and may be an important source of magnetospheric plasma.

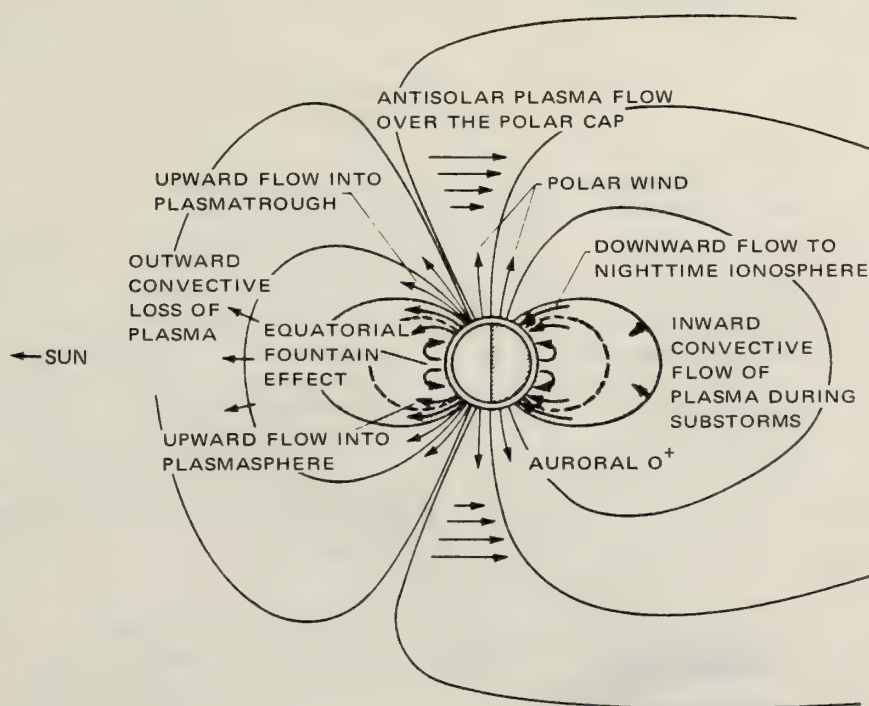


FIGURE 4 TRANSPORT PROCESSES THAT REDISTRIBUTE PLASMA IN THE IONOSPHERE-MAGNETOSPHERE SYSTEM (after Burch, 1977).

In the polar cap the lower plasma pressure on open field lines results in an upward flow referred to as the polar wind. At altitudes of a few thousand kilometers  $\text{H}^+$  and  $\text{He}^+$  flow out of the ionosphere with velocities of several kilometers per second, accompanied by an equal flux of electrons.

In the auroral zone there is a flow of energetic electrons and protons into the ionosphere. On the nightside they originate in the earth's plasma sheet and magnetotail. On the dayside they enter the ionosphere directly from the magnetosheath through the plasma cleft. This flow of plasma into the ionosphere results in the outflow of secondary electrons produced by atmospheric collisions. Also, the precipitating plasma generally carries a net negative charge into the ionosphere which must be balanced by a return current (see Figure 5). This return current is generally thought to be carried by upward-streaming thermal ionospheric electrons. The parallel electric field that accelerates auroral electrons downward also accelerates atmospheric ions ( $O^+$ ) upwards out of the ionosphere and into the magnetosphere. Satellite observations of ion composition in auroral arc flux tubes at altitudes of  $\sim 1 R_E$  consistently show upward fluxes of  $O^+$  ions. This means that the ionosphere, through the auroral arc parallel potential drop, supplies  $O^+$  ions to the magnetosphere. The precise implication of this observational fact is as yet unknown; however, the significant conclusion that the ionosphere is not a passive absorber of magnetospheric particles, but is an active and immediate participant in magnetospheric processes cannot be avoided.

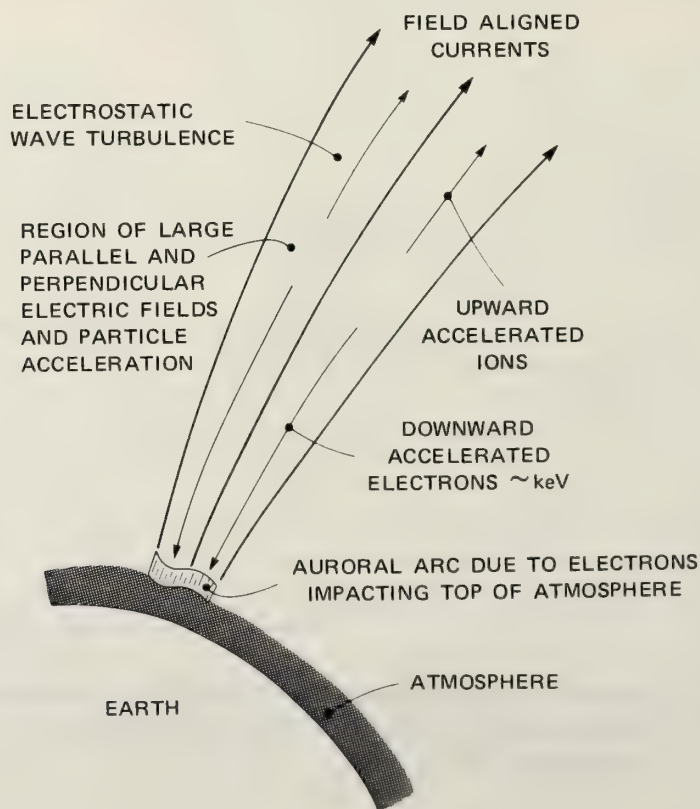


FIGURE 5 PICTORIAL REPRESENTATION OF THE FIELD-ALIGNED CURRENTS, PARALLEL ELECTRIC-FIELD REGION, AND OTHER PHENOMENA ASSOCIATED WITH AN AURORAL ARC (after Space Science Board, 1978).



Equatorward of the auroral zone, the geomagnetic field lines are similar to those of a dipole and can confine the ionospheric plasma. It has been surmised for a long time that this region of relatively dense plasma, known as the plasmasphere, is basically an extension of the topside F-region. Direct evidence, independent of the question of density distribution and composition, can be found in observations of thermal plasma flow at low latitudes by incoherent-scatter radar techniques. The thermal plasma that flows from one hemisphere to the other along magnetic flux tubes is observed to couple topside ionospheres which are under separate day-night and winter-summer situations. Although such plasma flow is probably not a major means of establishing equilibrium between the ionosphere and the plasmasphere, it is a very important means of transmitting the plasma response between asymmetric conditions of conjugate ionospheres. Also, during disturbed conditions, observations and theories suggest that increased convection flow in the plasmapause region may sweep away some flux tubes of thermal plasma. The ionospheric response to such a sudden loss of plasma at the equatorial regions of the flux tube is likely to be an upward plasma flow. Therefore, thermal plasma flow is a likely ionospheric response to deviations from equilibrium brought about by regular variations, such as the diurnal and seasonal cycle, or by irregular impulses, such as geomagnetic storms.

### 3. IMPORTANCE TO FORECASTERS AND USERS

Although all of the interactions described in the previous section influence the configuration and dynamics of the magnetosphere and ionosphere, not all of them produce effects significant enough to be of user concern. In Table 1 we have listed the ionospheric and atmospheric manifestations of magnetospheric-ionospheric coupling that can produce an impact on the user. We have attempted to categorize the importance of the potential impact as being major, minor, or uncertain. The impact on the user of the more significant processes are discussed in detail in this section. The reports of other working groups in this proceeding also discuss many of these action phenomena.

#### 3.1 High-Latitudes

The impact that high-latitude effects of geomagnetic disturbances has upon environmental forecasting is, a major topic of concern to other working groups. Here, we will restrict our discussion to magnetosphere-ionosphere coupling only. Even with such a restriction, the complexity of the phenomena in the high-latitude region causes considerable overlap with respect to forecasting and usage in other areas. In order to maintain some unity in this regard, we shall classify by morphological features and note the relevance of ionosphere-magnetosphere coupling as we proceed.

Table 1

## IMPORTANCE TO FORECASTERS AND USERS

<u>Region</u>	<u>Process</u>	<u>User Relevance</u>	<u>Significance</u>
High-Latitude Ionosphere			
1. E-Region	EF (auroral particles)	Communications (RP, TI)	Major
2. D-Region	EF (REP, PCA)	Communications (RP)	Major
3. F-Region	EC (Convection)	Communications (RP)	Major
4. Topside F-Region	PT (Polar wind)	Communications (RP)	Minor
5. Ionospheric irregularities	EC (currents)	Communications (TI, RP)	Uncertain
6. Auroral electrojet	EF (conductivity); EC (field & current)	LL/PL/PL Magnetometer Surveys	Major Major
Low-Latitude Ionosphere			
7. Plasmasphere	PT, EF	Communications (RP)	Major
8. Ionospheric irregularities	EC (electric field)	Communications (TI, RP)	Uncertain
9. Equatorial electrojet	EC (electric field)	Magnetometer Surveys	Major
Neutral Atmosphere			
10. Thermosphere	EF (Joule heating, auroral particles)	Satellite Drag	Major
11. Troposphere	EF, EC	Weather/Climate	Uncertain

EF = energy flow

EC = electrical coupling

PT = plasma transport

RP = reflected propagation

TI = transionospheric propagation

LL/PL/PL = long-lines, power lines, and pipelines



### 3.1.1 High Latitude Ionization

High latitude structures such as the trough, the auroral oval, the cusp and the polar cap result from the interaction between the solar wind and the magnetosphere. However, their location, size, amplitude and temporal development in the ionosphere are the result of magnetosphere-ionosphere interactions.

The ionospheric F-region trough marks the boundary between the low density polar ionosphere and the denser mid-latitude ionosphere. The formation of the F-region trough, similar to the formation of the plasmopause, is necessarily a feature of dynamical magnetosphere-ionosphere interactions, although the exact relationship between these two very similar plasma "boundaries" is still not understood. The F-region density gradients in the trough region play an important role in allowing intercontinental communication via HF, although present usage of these propagation strategies are by trial and error. Thus, predictive capability of the characteristics of the trough may be an effective means of cost reduction in user systems, rendering acquisition of sophisticated systems unnecessary. Although the existence of the trough has been known for a long time and the use of it is not new, precise scientific understanding of its relationships with its controlling features in the magnetosphere, such as the plasmopause and the boundary between "closed" and "open" field lines, are not known. Scientific modelling of trough features in the F-region seems to be needed as a prerequisite to predictive capability.

The particle and field dynamics in the auroral oval and polar cusp during geomagnetically disturbed periods are so complex that it would not be appropriate to enumerate them here. The key magnetosphere-ionosphere interaction that affects forecasting and usage is that in these regions the disturbance energy is injected into the ionosphere and neutral atmosphere. Hot plasma injection into the auroral zone from the plasma sheet and ionospheric currents driven by the convection electric field constitute the primary energy input in the auroral oval. Direct solar wind particle access in the cusp on the dayside is yet another mechanism of energy injection into the high-latitude ionosphere. Ionospheric disturbances in the high-latitude regions are of major concern to a number of communication and ranging systems. The size and location of the auroral oval is fairly well monitored by both ground-based and space-based systems. Major advances in the understanding of geomagnetic storms have led to construction of semi-physical predictive indices based on solar wind characteristics. So, operational predictive capability in terms of warnings of disturbances is presently available in principle. However, in order to achieve predictive capability in terms of the intensity level and morphology of the ionospheric storm, physical understanding of how this disturbance energy is coupled into the ionosphere is required.

Poleward of the auroral oval is a region of low ionospheric density, referred to as the polar cap, which is presumably due to upward evacuation by the polar wind and the absence of strong ionization sources. The physics of very large F-region density fluctuations in the polar cap region, known as polar troughs, is not known.

The complex high-latitude magnetospheric particle and field dynamics leads to morphological features in the high-latitude ionosphere through energy coupling mechanisms. The ionospheric state in this region affects primarily radio communication and ranging systems. At present, we have predictive capability to the extent of warning, but not to the extent of predictive modelling.



### 3.1.2 Substorm Dynamics

The spatial distribution and temporal variation of auroral ionization during substorms is of major importance to ionospheric users and forecasters. Operationally, individual substorms have a major impact in terms of communication links and ranging applications using the ionospheric D and E regions. The substorm is identified as an individual event of auroral particle injection and auroral electric field intensification. During the main phase of a geomagnetic storm, substorms occur very frequently, but substorms do not occur only during a magnetic storm. The time scales associated with substorm features is so short and the occurrence characteristics are so sporadic, that it is not known whether substorms can be predicted with any accuracy greater than that of a general warning associated with magnetic storms. In the last few years, as a result of satellite, rocket and radar observations, major advances in understanding the electrodynamics of magnetosphere-ionosphere coupling during substorms have been made. Needless to say, physical understanding of the magnetosphere-ionosphere interactions in the aurora is a precursor to the development of semi-physical predictive methods.

### 3.1.3 Thermospheric Heating

Energy input in the form of particle precipitation and Joule dissipation of ionospheric electric fields is an important source of heat for the thermosphere. During disturbed periods, this source of thermospheric heating from magnetosphere-thermosphere coupling is typically of the same magnitude as the heat source due to EUV absorption at high latitudes. As one may expect, thermospheric heating at the high-latitude region causes the thermosphere to expand upwards and equatorwards driving thermospheric winds in the process. Therefore, at a given altitude, the thermospheric density is increased and this increased density, in turn, implies increased drag for low-altitude satellites (Straus and Hickman, 1979). For a number of satellite systems, prediction and monitoring of satellite ephemerides are required; hence, the thermospheric heating aspects of magnetosphere-thermosphere coupling are of major importance to low-altitude satellite ephemeris predictions. Low altitude satellites typically have orbital periods of about 1.5 hours. On the other hand, substorm injection events also have similar lifetimes. With such similar time scales, it is expected that thermospheric drag effects on satellite ephemeris determination may be significant, even affecting individual revolutions.

## 3.2 Low- and Mid-Latitudes

At low and mid-latitudes, particle energy input into the ionosphere consists primarily of ring current particle energy precipitation. The response of the F-region ionosphere to geomagnetic disturbances is, however, not solely dependent on the energetic components of particle energy precipitation. Recent satellite observations of thermal plasma temperature at the equatorial regions of the plasmasphere indicate large increases ( $> 10^4$  K) during disturbed periods. Since thermal plasma in this region is directly connected to the ionosphere by observed field-aligned plasma flow, this energy transport into the ionosphere may perhaps be as important as direct energetic particle precipitation.

Magnetospheric VLF communication links depend critically on the structure of the topside ionosphere and plasmasphere. Energy coupling and echo properties of VLF ducts are affected by the altitude of the duct end-point (the  $H^+/O^+$  transition altitude). Because VLF signals are likely to be ducted, the requirements for VLF link predictions is probably more stringent than simply predicting an average plasmaspheric density. Except for isolated attempts at reconstructing VLF ray paths, there is at present neither capability nor attempt to predict VLF communication links. To be sure, systematic use of magnetospheric VLF communication will become a reality; thus, allowing the prediction community time to consider this difficult problem is a present concern.

Some 15% of the total-electron-content (TEC) for ground-to-satellite radio links are accounted for by the plasmasphere and topside ionosphere. While present day satellite communication systems are not severely affected by the topside ionosphere and plasmasphere, the plasmaspheric TEC directly impacts the accuracy of active and passive ranging systems. At present, average time delay corrections are applied for such systems, but predictive capability is non-existent.

There is increasing evidence that man-made sources of VLF power lines and transmitters may precipitate energetic particles from the radiation belts. Although no usage is to be derived from such an effect, quantitative assessment of this environmental impact may require predictive capability similar to that just described.

#### 4. MONITORING AND PREDICTION TECHNIQUES

##### 4.1 Present-Day Monitoring Methods

Monitoring techniques are described in several of the contributed papers to this proceedings (e.g., Thompson and Secan, 1979) and in most of the other working group reports. Here we will consider techniques relevant to phenomena that involve magnetospheric-ionospheric interactions.

##### 4.1.1 High Latitudes

The real-time monitoring of ionospheric-magnetospheric interactions is generally noted by observing the effects on high-latitude ionospheric morphology. These effects can be observed through a number of ground-based and satellite observational systems. These include vertical sounders, riometers, Faraday polarimeters, auroral radars, incoherent-scatter radars, ion-electron plasma probes, magnetometers, optical auroral imagers, and charged-particle detectors.

Indirect monitoring of the auroral electrojet position and intensity can be made using auroral radars and magnetometer networks. These data are not routinely used for this purpose except in a very subjective fashion. Another cruder estimate is obtained from optical aurora and particle precipitation data.

No real-time monitoring of thermospheric heating is made. The indirect effects are noted by satellite position errors but these data cannot be considered as a useful measurement of thermospheric heating.



#### 4.1.2 Plasmasphere

Monitoring the plasmasphere-ionosphere system requires not only measurement of the thermal plasma parameters (density, composition and temperature), but also the state of energetic particles and electric fields which control the energy input into the system during disturbances. Thermal plasma is monitored primarily by satellite instruments, although at selected locations incoherent backscatter radar is used to measure thermal plasma parameters at low altitudes. Under some assumptions, VLF transmission characteristics can be an effective monitor of plasmaspheric density in the equatorial region.

Satellite instruments, such as those onboard the ISEE, GOES, and S 3-3, have acquired vast amounts of plasmaspheric data; however, a semi-theoretical model is required to sort out the space and time effects before these data can be used as effective plasmasphere monitors. Incoherent backscatter radar yields data on density and plasma flow without the problem of mixing space-time effects; however, it is limited by spatial scope and operational availability. VLF transmission characteristics also suffer from limited spatial scope and lack of operational control. A merging of the three types of monitoring data in a forecast center scenario may overcome their individual shortcomings, but it has not yet been attempted.

### 4.2 Prediction and Specification Techniques

#### 4.2.1 High Latitudes

There are many specialized models of the high latitude ionosphere (e.g., Elkins and Rush, 1973; Vondrak et al; 1977), but most have serious deficiencies or are restricted to only certain ionospheric layers. A more comprehensive model (known as Polar-Ionospheric) is presently under development at the Air Force Global Weather Central (AFGWC). This model is essentially an extension of the AFGWC 4-D model (Tascione et al., 1979) that combines the best features of all the existing high latitude models. The model will specify ionospheric conditions in the polar cap, diffuse aurora, auroral trough, high-latitude irregularity zones, and the mid-latitude ionosphere and will be able to provide gridded electron density altitude profiles from 90 to 500 km. The inputs will be  $K_p$ ,  $Q_e$  and existing ground based vertical sounder data. The model is scheduled for operation in October 1979.

At AFGWC auroral substorm activity is specified from DMSP optical and particle precipitation data. These data are used to calculate two indices of auroral activity,  $Q_e$  and  $C_a$  (Pike, 1975). The effective Q index,  $Q_e$ , is derived after determination of the equatorial edge of the diffuse aurora from either optical or particle precipitation data. The presence of an auroral substorm can also be inferred by using auroral radar and/or geomagnetic observations. A simple forecast of substorm onset is made by using a technique described by Baker et al., (1979). Indirect specification of the auroral electrojet position is made using the DMSP optical aurora and particle precipitation data. The auroral radar data could also be used in this fashion but is not due to operational problems.



#### 4.2.2 Mid-Latitudes

Probably the most complete ionospheric specification model operational at the present time is the AFGWC 4-D ionospheric model (Tascione et al., 1979). This spectral analysis model uses as input a wide variety of ground based and satellite observations. It computes electron density profiles for any latitude, longitude, and time. However, due to spatially-limited data inputs and to model deficiencies in the handling of large ionospheric gradients, it is considered to be basically a mid-latitude model.

The Polar-Ionospheric and 4-D models are both basically specification models and, therefore, have no time scales of prediction. The 4-D ionospheric model has the following accuracy in the northern hemispheres:

foF2	$\pm .5$ MHz
Muf/FOT	$\pm 3.0$ MHz
TEC	$\pm 25\%$
Electron Density Profile	$\pm 25\%$ RMS error

#### 4.2.3 Thermosphere

At present, physical modelling of auroral heating of the high-latitude thermosphere is available, although operational use of such models for prediction and forecast purposes has yet to be accomplished. The time dependence of the heat input becomes a significant feature of satellite drag effects on individual orbits; hence, we would expect future predictive models to include this important feature. It must be pointed out that high-latitude heating is only one of many factors affecting thermospheric density. Therefore, predictive models of thermospheric density changes due to magnetospheric energy input are predictive only in the sense of reduction of ephemeris prediction error residue. At present, it is not known exactly what percentage reduction can be accomplished by an accurate specification of the effects of high-latitude heating.

### 5. FUTURE PROGRESS-GOALS AND EXPECTATIONS

New sophisticated satellite and ground-based instrumentation currently in the planning and development stages will provide magnetospheric and ionospheric data that is more-detailed, of higher resolution, and geographically more extensive than is currently available. These advanced programs are the result of increases in our understanding of magnetospheric and ionospheric phenomena. However, in many instances progress in data gathering capability is more rapid than can be incorporated into specific models and prediction schemes, because many of the current prediction schemes are relatively crude and antiquated. Although current knowledge must guide the development of data-gathering platforms for future-generation prediction models, immediate needs may be satisfied through advances by other means. Significant progress would result from increased rapport between the experimentalist and the user as well as increased

use of the currently available, yet relatively unused, wealth of geophysical data. Thus, a strong effort should be made to use the currently available data rather than waiting for the results of new, exotic, data-gathering programs.

To improve existing forecasting programs, existing satellite measurements of solar wind parameters (such as  $B_z$ ,  $B_x$ , and  $V$ ) should be used to predict auroral oval characteristics, total magnetospheric energy input, and current system morphology.

Satellite programs that measure magnetospheric parameters with greater precision and at more suitable locations will clarify the relationships between magnetospheric changes and ionospheric response.

High altitude satellites using optical or X-ray imaging techniques will yield real time global pictures of the ionospheric response to the entry of particles from the magnetosphere and these will greatly aid the forecasting of detailed changes in structure, size and location of enhanced electron density regions in the ionosphere. Because energetic particle precipitation optically excites the atmosphere to produce the aurora, the spatial extent and magnitude of this energy input can be determined optically. However, the flow of current in the ionosphere does not have an analogous signature and, moreover, it cannot be presumed that these currents necessarily flow where the visual aurora is located. Indeed, there is evidence that the regions of large dissipative current flow are not spatially co-incident with the brightest features of the visual aurora.

In many forecast areas, such as oval size, internal dynamics, particle distributions, and the influence on neutral composition, our future ability to specify adequately the current state and to predict a few hours ahead seems quite bright. Longer-term predictions made prior to a magnetospheric storm will require a greater understanding of its relation to solar activity and the identification of interplanetary parameters which are precursors of its onset. Such developments seem to be many years in the future.

Attainment of a reliable forecast capability can be accomplished by pursuing several intermediate objectives, each of which is useful to the forecast user. One possible approach, in which the objectives are ordered in increasing complexity, is as follows.

1. Monitor the instantaneous spatial pattern of energy input to the atmosphere from the magnetosphere, together with an estimate of total energy input by particle precipitation and Joule heating. Monitoring of these parameters will establish a data base from which one may be able to generate empirical relations between the geophysical disturbance level and specific ionospheric effects. Also, as discussed in Section IV, a quantitative knowledge of geophysical activity can be used to compute some immediate ionospheric consequences. For example, the particle energy input may yield a specification of the electron density profiles in the D, E, and lower F-regions.

2. Use the present instantaneous pattern of energy input to estimate future consequences. For example, the current state of thermospheric heating could be used to forecast the probable location and duration of neutral density enhancements and traveling ionospheric disturbances. The spatial distribution of F-region ionization at high latitudes could, in principle, be computed from measurements of the electric field convection pattern and the ionization production rate.



3. Use the measurements of recent history of energy deposition patterns and magnetospheric particle distributions to predict the magnetospheric response to subsequent events. Either low-altitude or high-altitude satellites can be used to monitor the number density and temperature (characteristic energy) of the plasma sheet reservoir. With this information it may be possible to predict how much energy will be transferred to the ionosphere by future magnetospheric perturbations. Thus, the amplitude of a potential magnetospheric disturbance could be predicted without necessarily being able to predict the exact time of occurrence.

4. By monitoring the state of the magnetosphere, and understanding the mutual coupling between the magnetosphere and ionosphere, predict when a disturbance is likely to occur as well as its intensity and probable effects upon the atmosphere and ionosphere. This objective is, of course, the ultimate one. The major requirement is a full understanding of energy storage release in the magnetosphere-ionosphere system, the transfer of this energy to the atmosphere, and all subsequent effects, including possible feedback into the magnetosphere.

## 6. RECOMMENDATIONS

In order to continue the progress that has been made in the application of magnetosphere-ionosphere interaction effects to user needs, we recommend the following:

### 6.1 Increased Use of Existing Capabilities

1. Increased emphasis and effort in data analysis. A surprisingly small amount of experimental data is ever analyzed. More emphasis and money needs to be placed upon data analysis by the funding agencies.

2. The data needs to be better organized for use by both scientists and the forecaster/user. This effort as well as improved distribution is the responsibility of the various data centers. It will, however, require considerable support and assistance from the scientists and forecasters/users. Selected data types should probably receive first consideration.

3. Develop means to extract simple operationally usable data from existing and planned satellite and ground-based systems. This data must then be provided to the forecasters in real time. This should greatly benefit all.

4. It is strongly recommended that increased forecaster/user-scientist interaction take place. The gap between these two groups is ever-widening as new knowledge is developed faster than it can be applied. Further, it is noted that this interaction must take place in a continuous manner and that one meeting does not fulfill this need.

### 6.2 Development of Improved Monitoring Techniques

The following parameters need to be monitored on a routine basis:

1. Global auroral activity. By continuous satellite imaging in many wavelengths (visible, x-ray, UV) it may be possible to deduce E-region



electron density profiles over the locations of auroral particle precipitation.

2. Particle precipitation and energy input at high latitudes, as well as the plasma distribution functions in the magnetosphere, particularly the plasma sheet.

3. Electric fields, electric currents, and conductivity distributions at high latitudes.

4. Plasma density, temperature, and velocity in the plasmasphere.

### 6.3 User Agency Support of Ionospheric and Magnetospheric Research

The user agencies should recognize that "scientific" models must precede operational models by identifying key relationships, the best parameters to monitor, and establishing indices. They also serve as the framework upon which simplified operation models can be based. User agency support should be given to the development of such scientific models, despite their having the appearance of basic research.

## 7. MAGNETOSPHERE-IONOSPHERE INTERACTIONS WORKING GROUP

R. R. Vondrak, Chairman  
Radio Physics Laboratory  
SRI International  
Menlo Park, CA 94025

V. G. Patterson  
Technical Services Division  
Air Force Global Weather Central  
Offutt AFB, NE 68113

Y. T. Chiu  
The Aerospace Corporation  
P. O. Box 92957  
Los Angeles, CA 90009

G. J. Romick  
Geophysical Institute  
University of Alaska  
College, AK 99701

D. S. Evans  
NOAA Research Laboratories  
Space Environment Laboratory  
Boulder, CO 80303

K. Stasiewicz  
Space Research Center  
Polish Academy of Sciences  
Ordonia 21, Warsaw, Poland

## 8. REFERENCES

- Akasofu, S.-I., "Recent Progress in Studies of DMSP Auroral Photographs," Space Sci. Rev., 19, 169, 1976.
- Akasofu, S.-I., Physics of Magnetospheric Substorms, D. Reidel, Dordrecht, Holland, 1977.
- Baker, D., P. Higbie, E. Hones, and R. Belian, "The Use of  $> 30$  keV Electron Anisotropies at  $6.6 R_e$  to Predict Magnetospheric Substorms," this proceedings, 1979.

- Boström, R., "Magnetosphere-Ionosphere Coupling," in Critical Problems of Magnetospheric Physics, National Academy of Sciences, Washington, 1972.
- Burch, J., "The Magnetosphere," in The Upper Atmosphere and Magnetosphere, National Academy of Sciences, 1977.
- Chiu, Y., Cornwall, and M. Schulz, "Auroral Magnetospheric-Ionospheric Coupling: A Brief Topical Review," this proceedings, 1979.
- Elkins, T. and C. Rush, "A Statistical Predictive Model of the Polar Ionosphere," in An Empirical Model of the Polar Ionosphere, T. Elkins, ed., AFCRL-TR-73-0331, Air Force Cambridge Research Laboratories, Hanscom AFB, Massachusetts, 1973.
- Kamide, Y., "Recent Progress in Observational Studies of Electric Fields and Currents in the Polar Ionosphere: A Review," Antartic Record, 63, 1979.
- Pike, C., "Defense Meteorological Satellite Program Auroral-Ionospheric Interpretation Guide," AFCRL-TR-75-0191, Air Force Geophysics Lab, Hanscom AFB, MA, 1975.
- Schunk, R. W., P. M. Banks, and W. J. Raitt, Effects of Electric Fields and Other Processes Upon the Nighttime High-Latitude F layer, J. Geophys. Res., 81, 3271-3282, 1976.
- Space Science Board, Space Plasma Physics: The Study of Solar-System Plasmas, National Academy of Sciences, Washington, D.C., 1978.
- Straus, J. and D. Hickman, "Predictability of Upper-Atmospheric Density and Composition," this proceedings, 1979.
- Tascione, T., T. Flattery, V. Patterson, J. Secan, and J. Taylor, "Ionospheric Modeling at Air Force Global Weather Central, this proceedings, 1979.
- Thompson, R., and J. Secan, "Geophysical Forecasting at AFGWC," this proceedings, 1979.
- Vondrak, R., G. Smith, V. Hatfield, R. Tsunoda, V. Frank, and P. Perreault, "Chatanika Model of the High-Latitude Ionosphere for Application to HF Propagation Prediction, RADC-TR-78-7, Rome Air Development Center, Hanscom AFB, MA, 1977.
- Wolf, R., "Ionosphere-Magnetosphere Coupling," Space Sci. Rev., 17, 537, 1975.

# AURORAL MAGNETOSPHERE-IONOSPHERE COUPLING: A BRIEF TOPICAL REVIEW

Y. T. Chiu, J. M. Cornwall<sup>\*</sup>, and Michael Schulz  
Space Sciences Laboratory  
The Aerospace Corporation  
P. O. Box 92957  
Los Angeles, CA 90009

<sup>\*</sup>Permanent address: Department of Physics  
University of California, Los Angeles, CA 90024

Auroral arcs result from the acceleration and precipitation of magnetospheric plasma in narrow regions characterized by strong electric fields both perpendicular and parallel to the earth's magnetic field. The various mechanisms that have been proposed for the origin of such strong electric fields are not mutually exclusive. Indeed, they are often complementary. Such mechanisms include: 1) electrostatic double layers; 2) double reverse shocks; 3) anomalous resistivity; 4) magnetic mirroring of hot plasma; and 5) mapping of the magnetospheric-convection electric field through an auroral discontinuity. Observations have not yet identified from among these mechanisms the one that is primarily responsible for the formation of auroral arcs.

## I. INTRODUCTION

The latter half of the nineteen seventies has witnessed a spectacular increase in the understanding of the detailed dynamics of the auroral arc. This happy circumstance is supported on the one hand by the advent of high-resolution observations by auroral satellites at altitudes  $\sim 1 R_E$ , such as those obtained by instruments on board the S3-3 satellite, and on the other by earlier rocket and radar backscatter observations at ionospheric altitudes. Since auroral processes are a major element in the coupling between the magnetosphere and the ionosphere, we undertake in this paper to review briefly what the new satellite observations imply and do not imply in the theoretical understanding of auroral dynamics in general and the coupling between the ionosphere and the magnetosphere in particular.

Because of the self-imposed restriction of brevity, the scope of this topical review is limited to those dynamical aspects of the aurora directly



related to the detailed processes of ionosphere-magnetosphere coupling. The very important optical observations of the aurora fall in a different category and are not included here. Global-scale morphology of ionospheric currents generated by auroral precipitation of energetic particles and the mapping of auroral electrostatic fields down to ionospheric heights are important consequences of the aurora; however, these effects are consequential rather than fundamental to the coupling between the ionosphere and the magnetosphere so we shall allude to them only briefly. In short, we limit our consideration to recent advances in the understanding of particle and field dynamics as they affect ionosphere-magnetosphere coupling in the aurora.

Further, since brevity is emphasized, works and authors referenced here are meant to be representative rather than exhaustive. We attempt to reference the latest work (circa 1978) representative of a particular area so that readers can find a more complete reference therein; therefore, works referenced in this paper do not necessarily carry with them any implication of special significance. Perhaps undue emphasis is placed on the implications of the S3-3 observations, which to some extent merely confirm what had been suspected from earlier rocket and satellite measurements; however, this is partially due to the timing of this review, falling at the point when the scope and coherence of the S3-3 observations have become clear to those of us who are interested in the theoretical aspects of ionosphere-magnetosphere coupling and possible predictions.

## II. OBSERVATIONS

Before the advent of high-resolution observations by auroral satellites, data on ionospheric and magnetospheric coupling depended on balloon, rocket, and radar observations which were necessarily episodic; nevertheless, the basic physical properties of auroral ionospheric currents and electric fields (e.g., Cloutier, 1971; Mozer and Manka, 1971; Vondrak et al., 1971), together with their relationship to high-latitude convection electric fields (e.g., Cauffman and Gurnett, 1972; Heppner, 1972) measured by satellites, have been established. Generally, these measurements have indicated that the substorm convection electric field in the magnetosphere drives perpendicular ionospheric currents consistent with ionospheric perpendicular electric fields (meridional and zonal) of the order of tens of millivolts/meter. In addition, Birkeland currents parallel to auroral magnetic field lines seem to have been observed (e.g., Armstrong and Zmuda, 1973). These low-altitude observations are primarily concerned with the morphology and large-scale processes of auroral substorms and are instrumental in emphasizing the importance of the electric field in auroral processes. They, however, have relatively little to say about the microscopic processes taking place in the auroral region.

Even as the large-scale auroral processes were being unravelled, certain microscopic features of auroral low-energy particle precipitations were being discovered. Frank and Ackerson (1971) noted that occasionally

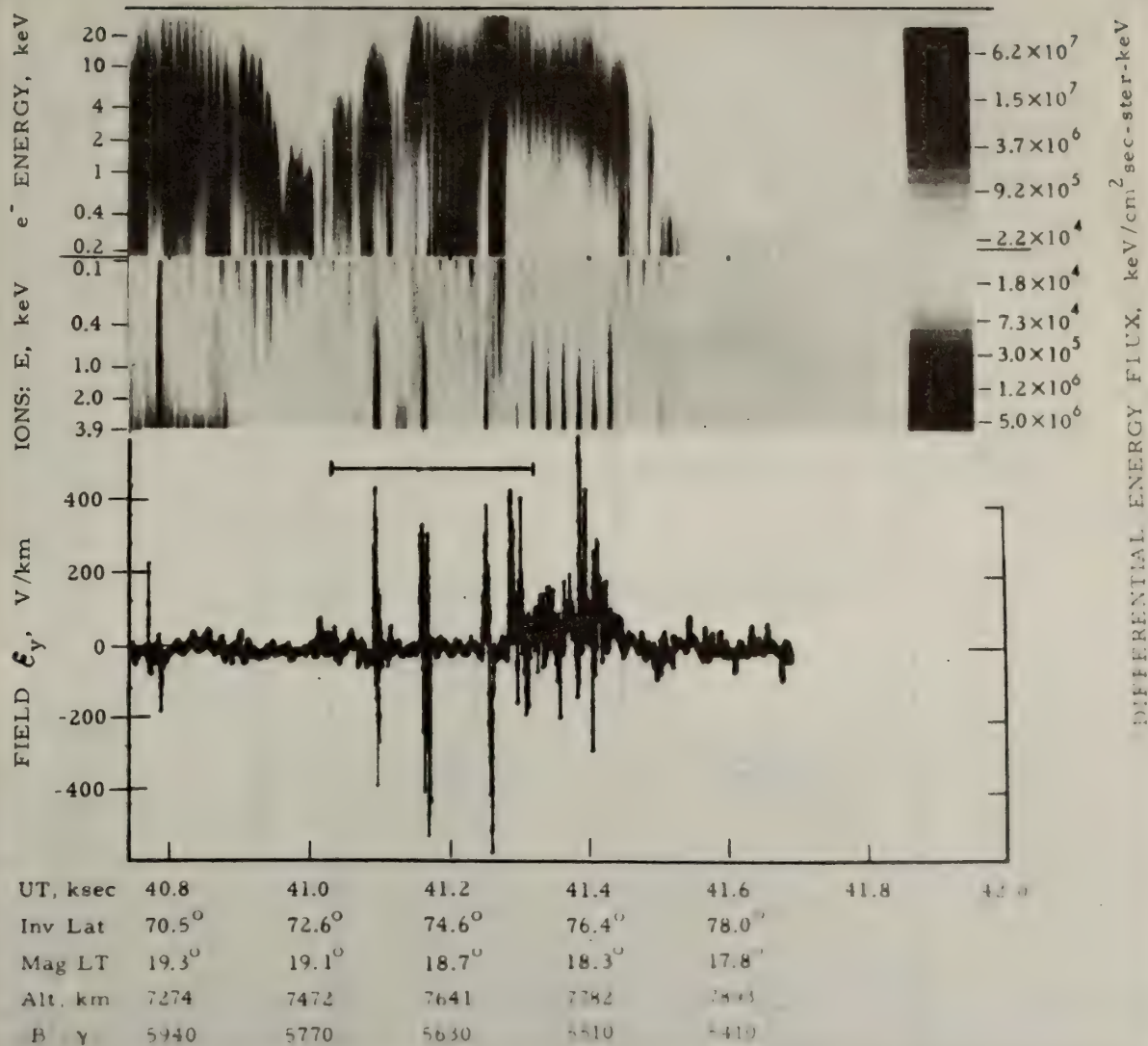
observations of low-energy (tens of keV) electron precipitation would show an "inverted-V" structure on an energy-time spectrum plot, i.e., the precipitating electron energy spectrum hardens and then softens as the Injun 5 satellite moves through the structure. Evans (1974; 1975) convincingly demonstrated that rocket measurements of auroral low-energy electrons indicated downward moving electron beams at keV energies, comparable to those of "inverted-V" structures. Further, by a careful study of electron backscatter from the atmosphere, Evans demonstrated that these auroral electron beams are indications of electric potential drops, along the magnetic field, existing between the equator and the ionosphere. At about the same time, observations of singly ionized energetic  $O^+$  ions in the magnetosphere (Shelley et al., 1972; Sharp et al., 1974) also gave indication that microscopic processes in the aurora couple the ionosphere with the magnetosphere.

These observations of "inverted-V" structures, of electron beams, and of  $O^+$  ions of probable ionospheric origin in the magnetosphere presage very interesting microscopic processes to be discovered in the auroral process in which the ionosphere plays an active rather than passive role. However, because of the episodic nature of rocket observations and because of the low resolution and low data rate of the early satellite observations, the scope of and inter-relationship between these phenomena were not understood until the launch of the polar-orbiting auroral satellite S3-3, which intercepts auroral field lines at altitudes up to  $\sim 8000$  km, precisely in the region where ionospheric and magnetospheric plasmas are expected to interact. Included in the S3-3 payload are instruments to measure electric fields (Mozer et al., 1977), low energy electrons (Mizera et al., 1976), energetic ions (Shelley et al., 1976), and plasma waves (Kintner et al., 1978). While a co-ordinated data-analysis program among the various auroral measurements is presently being pursued, the separately reduced data have already yielded a coherent picture of the microscopic auroral processes in which magnetosphere-ionosphere coupling plays a central role.

The S3-3 observations not only confirmed Evans' observations of downward-moving field-aligned electron beam at keV energies but also revealed the existence of upward-moving ion beams aligned with the magnetic field in "inverted-V" structures. This certainly indicates that the three phenomena are intimately related, but more importantly the S3-3 particle observations leave little doubt that an electric potential drop of several to tens of kilovolts, aligned with the magnetic field, exists between the ionosphere and the magnetospheric equator. Electrostatic field measurements also indicate paired regions of oppositely directed perpendicular electric fields, with latitudinal scale lengths of some 50 km, reflecting a negative space-charge region presumably associated with downward-streaming electrons. Figure 1, which is a composite of particle and electric field data illustrates the above points. For further emphasis, Figure 2 shows an enlarged view of the perpendicular electric field data for the time period marked by the brace in the middle of Figure 1. A crucial, but seldom emphasized, feature which is brought out by the high sensitivity and high resolution of the S3-3 measurements is that the above correlated features are observed at the auroral zone pass after pass at all



S3-3, 29 JULY 1976





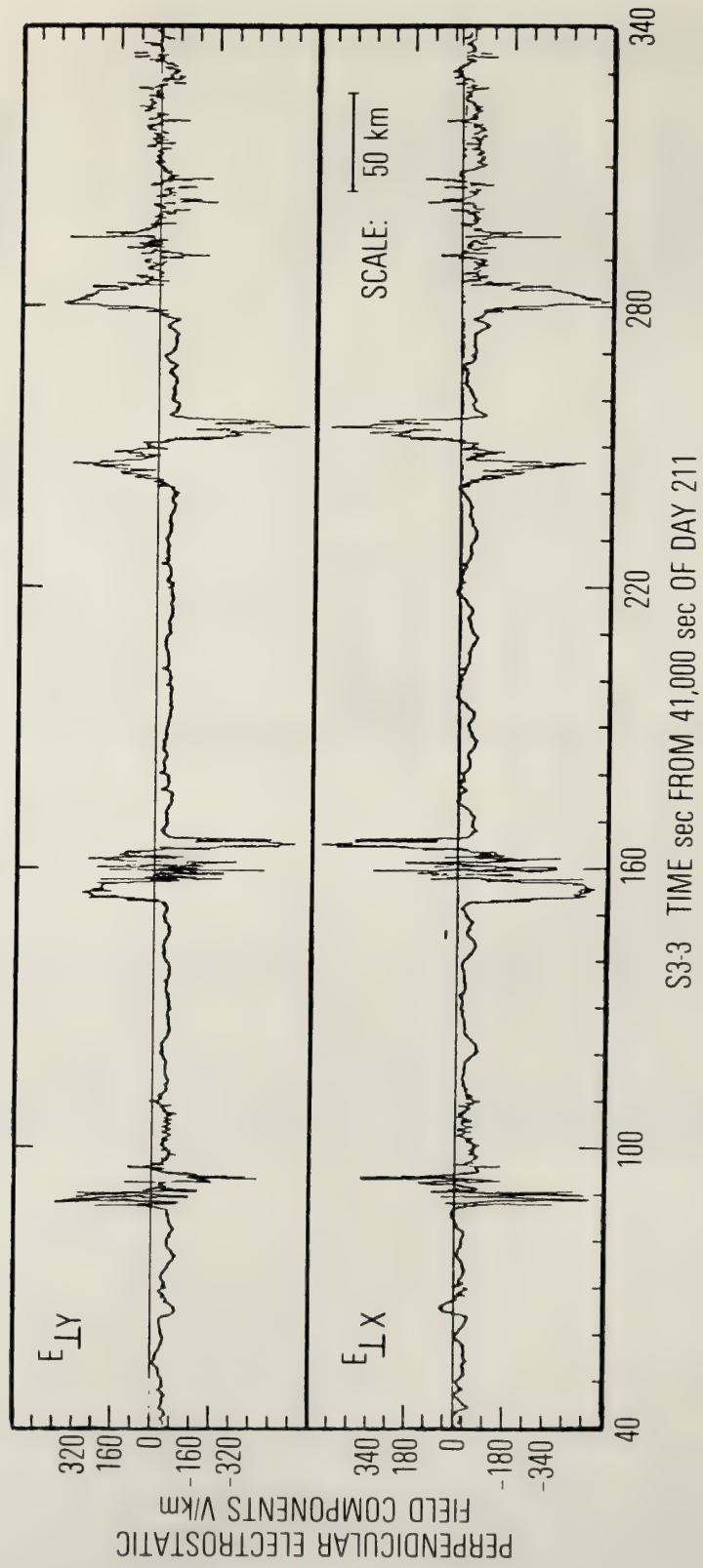


Figure 2.

satellite altitudes. In other words, these are fundamental features of the aurora rather than episodic curiosities. Further, plasma waves are observed to be associated with particle beams.

Aside from measurements which give, for the first time, support to a simple electrostatic picture of auroral microscopic processes, the S3-3 observations also reveal a whole class of new phenomena. The most outstanding among these are: 1) observations of "conical" beams, i.e., intense ion fluxes with pitch-angles concentrated on a cone about the direction of the magnetic field, with relatively little ion flux along the magnetic field; 2) upward-going field-aligned electron beams, and 3) downward field-aligned ion beams which are more diffuse than the upward field-aligned ion beams. The signature of conical beams is a bifurcated trace on the ion spectrograms; one can see several examples in Figure 1. Such beams probably result from wave-particle interactions with the basic auroral particle beams, while the downgoing ions and upgoing electrons, observed at lower altitudes, are a signature of the return current driven by electrostatic processes in the ionosphere. The S3-3 auroral observations have opened new vistas in the theoretical study of auroral processes.

### III. THEORETICAL INTERPRETATIONS

The key theoretical issue concerning the interpretation of the S3-3 observations really involves the electrodynamics of the auroral arc itself. A key fact which must be recognized is that a magnetic-field-aligned electrostatic potential difference of kilovolt magnitude exists between the ionosphere and the equator. In many respects, this feature has been anticipated in a number of theoretical considerations based on earlier observations. With regard to ionosphere-magnetosphere coupling, however, the crucial question is whether the mechanism for the buildup of such a field-aligned potential drop involves the ionosphere, for there is no doubt that the energy source of the aurora is derived from hot magnetospheric plasmas which are injected by substorm processes onto the auroral field lines. Some aspects of these theories of auroral field-aligned potential drop have been reviewed (Shawhan et al., 1978; Hudson and Mozer, 1978), but our discussions will be primarily concerned with the ionosphere-magnetosphere coupling aspects of these theories.

Theories of auroral processes involving magnetic field-aligned electrostatic potential differences can be roughly classified into five categories, although they are not mutually exclusive. These are: 1) double layer, 2) oblique electrostatic shock, 3) anomalous resistivity, 4) magnetic mirroring effects of differential pitch-angle anisotropy between ions and electrons, and 5) downward mapping of convection electric field discontinuities. These categories invoke theoretical arguments of varying degrees of sophistication and believability to show that kilovolt electrostatic potential drops may be produced in various assumed plasma distributions. The double-layer model is sharply differentiated from the others by its prediction of the scale length with which the total field-



aligned potential difference is distributed, i.e., the magnitude of the parallel electric field.

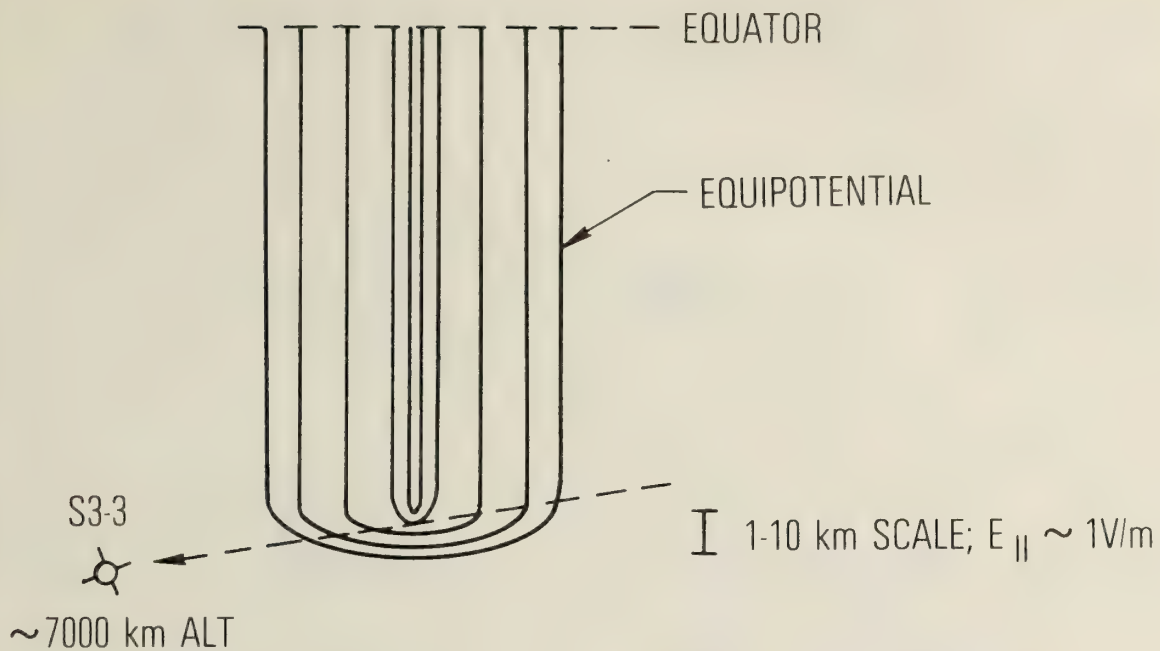
For the most part, these mechanisms have been considered in isolation of each other and of the ionosphere, not because physicists believe that it should be so, but because it is difficult to treat the couplings. In fact, a correct theoretical treatment of auroral phenomena will without doubt merge several of these mechanisms with each other and with ionospheric physics. It is unfortunate that much of the recent literature on auroral mechanisms pays so little attention to coupling with the ionosphere; some exceptions (with two of which the authors are connected) exist, though. As a general rule, the ionosphere couples neighboring field lines and allows for predictions of latitudinal structure and scale lengths. There is no such coupling in the individual mechanisms mentioned above (except that oblique shocks have an arbitrary structure which crosses field lines), so none can explain arc structure without going beyond the given mechanism. Our discussion begins with the traditional view of these mechanisms in isolation, then proceeds to a brief discussion of coupling schemes.

The double layer (Block, 1975; Shawhan et al., 1978) is a boundary layer between unmagnetized cold plasma on one side and hot plasma on the other. The potential drop across the layer is alleged to be  $\geq kT_e/|e|$  and the layer thickness is of the order of several Debye lengths ( $\sim 10$  km); thus, the parallel electric field in double layers must be  $\sim (0.1 - 1)$  V/m. A current-driven instability is usually invoked as the formation mechanism of double layers, which requires a field-aligned current greater than a certain threshold value. If potential drops inferred by electron beam observations at S3-3 altitudes as high as  $\sim 8000$  km are all due to double layers above the satellite, then evidently the ionosphere does not seem to be a factor in double layer formation either. Frequently, based on observations of both electron and ion beams on S3-3, one may infer that potential drops exist both above and below the satellite (Mizera and Fennell, 1977; Croley et al., 1978). Since it is highly improbable that the satellite just happened to pass through within the double layer thickness, such frequent occurrences seem to require more than one double layer to be formed on the same field line. Theories of double layer formation are mathematically difficult, even for very simple plasma distributions (Montgomery and Joyce, 1969), and a quantitative theory has yet to be developed for auroral plasmas. Even supposing that the theory is finally developed, and that difficulties of interpreting satellite data in terms of double layers can be overcome, there is one fundamental problem with double-layer models. They do not account for the influence of the earth's magnetic field, which -- except for the unrealistic case of a double layer exactly perpendicular to a magnetic field line -- is unwarranted, as we point out below. With regard to our main subject of magnetosphere-ionosphere coupling, double layers are so thin that they are almost completely decoupled from the ionosphere themselves. Furthermore, they tend to decouple the ionosphere from the magnetosphere above the double layer by effectively short-circuiting the magnetospheric electrical structure well above the ionosphere as indicated on Figure 3. In such a model, the ionosphere interacts little with the magnetosphere.

Oblique electrostatic shocks (Swift, 1975; 1976; Kan, 1975) are similar to double layers except that they recognize the influence of the



# DOUBLE LAYER



MAGNETOSPHERIC ELECTROSTATIC  
FIELD SHORT-CIRCUITED

----- IONOSPHERE ~100 km ALT

Figure 3.

magnetic field and consider that the shock normal is at an arbitrary angle  $\alpha$  to the magnetic field direction. For  $\alpha \neq 0$  the shock thickness  $\ell$  is measured in units of the ion gyroradius of a few km, a typical cross-field scale being some 15-20 gyroradii ( $\sim 100$  km). The field-aligned scale length is  $\ell/\cos \alpha$  which can be quite extensive if the shock normal angle  $\alpha$  approaches  $\pi/2$ . Swift has shown that self-consistent oblique shock solutions can be obtained with simple plasma distributions not unlike auroral conditions. A schematic illustration of Swift's double reverse electrostatic shock is shown in Figure 4. In addition to the fact that solutions of Poisson's equation have been obtained for semi-realistic plasma distributions in a homogeneous magnetic field, the oblique shock geometry has certain advantages over the current-driven double layer in regard to the interpretation of S3-3 data, even though the theory was conceived prior to S3-3. This is because the field-aligned scale length  $\ell/\cos \alpha$  can be chosen to be of the order of  $1-2 R_e$  so that only one shock (or a pair of double reverse shocks) need be invoked to explain the existence of potential drops above and below the satellite. It is, of course, a disadvantage that the theory as developed by Swift does not predict  $\alpha$ , or equivalently the cross-field scale length. As we discuss later, this scale length can be estimated by incorporating ionospheric physics. An oblique shock with parallel scale length of  $\geq 1 R_e$  is almost certainly strongly coupled to the ionosphere, which at the very least supplies important boundary conditions for the shock. The ionosphere and the magnetosphere tend to be strongly coupled as well, if only because the shock is so extended along the magnetic field.

A third mechanism by which a magnetic field-aligned electric potential drop can allegedly be generated is anomalous resistivity in the field-aligned direction (Hudson et al., 1978). Such anomalous resistivity may be due to a large number of possible modes of AC electric-field turbulence in the auroral plasma (e.g., Kindel and Kennel, 1971; Papadopoulos and Coffey, 1975). Hudson et al. (1978) estimated that turbulent electric fields in the electrostatic ion cyclotron mode with amplitudes  $\sim 50$  mV/m may yield sufficient anomalous resistivity to generate parallel electrostatic (DC) fields of  $\sim 1$  mV/m. However, it is not clear how the largely perpendicular AC fields can affect parallel electron currents (and their resistivity). One feature common to oblique-shock models and anomalous-resistivity models is that the potential smoothly varies over a scale of  $\sim 1 R_e$  extension in order to accommodate potential drops of  $\sim (1-10)$  kilovolts. This is schematically illustrated in Figure 5. It must be noted that the question of how such an extensive region of turbulence can be maintained at a high level ( $\sim 50$  mV/m AC), in the presence of non-linear stabilizing effects such as ion heating, must be addressed. A second feature of anomalous resistivity is that, unlike oblique shock models, there is little apparent relationship between the parallel and perpendicular electrostatic fields. In regard to ionosphere-magnetosphere coupling, the ionosphere plays a major role in at least one consideration of current-driven instabilities (Kindel and Kennel, 1971) since the effects of very weak ion-neutral and electron-neutral collisions,  $\sim 2 \times 10^{-2}$  of the cyclotron frequency, are stabilizing, as are the effects of ion heating. It must be said, however, that hydrogen ion-cyclotron waves are measured (Kintner et al., 1978) and there is little doubt that these waves will turn out to play some role in the dynamics of the auroral beams.

## DOUBLE REVERSE SHOCK

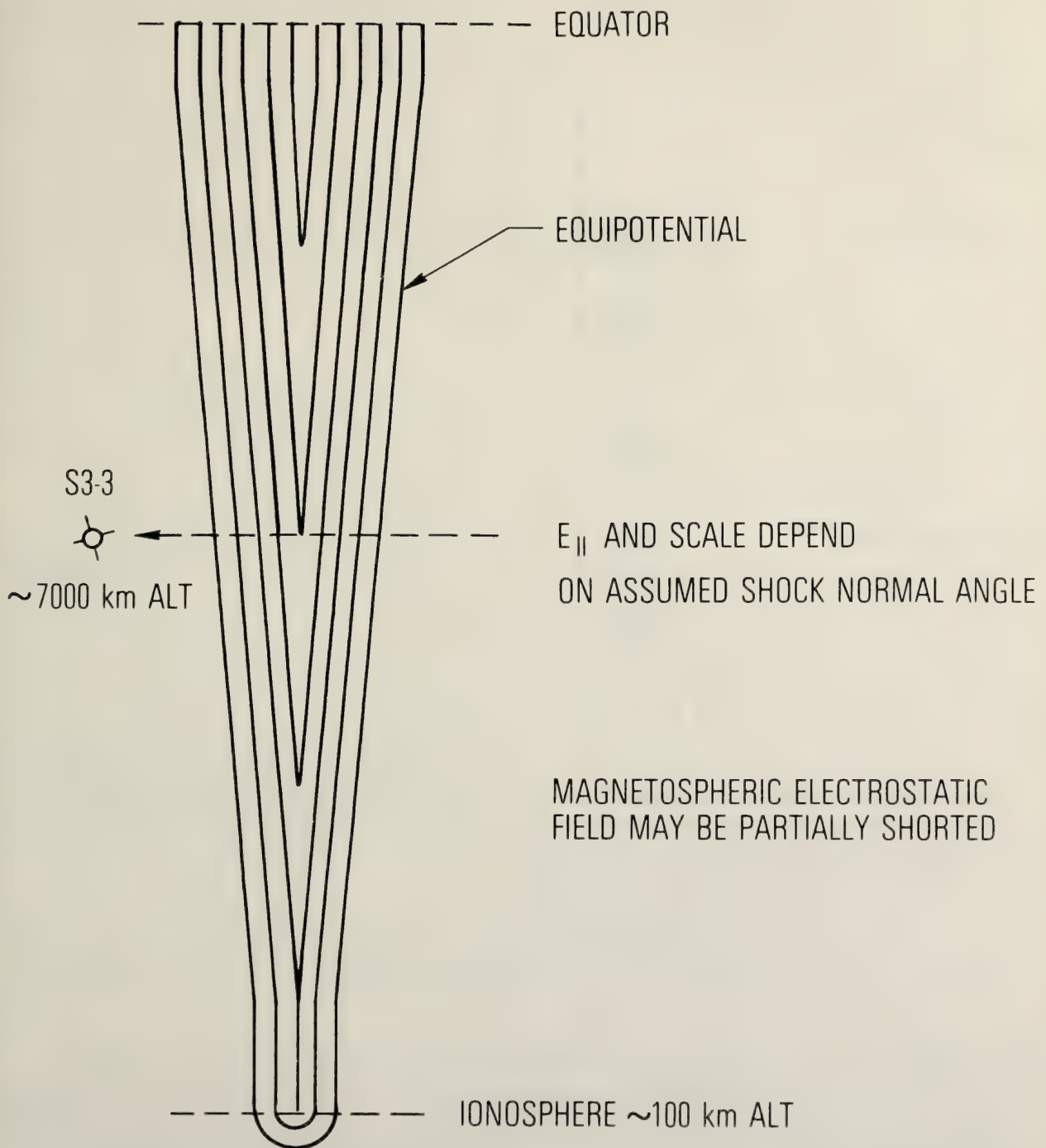


Figure 4.



# ANOMALOUS RESISTIVITY

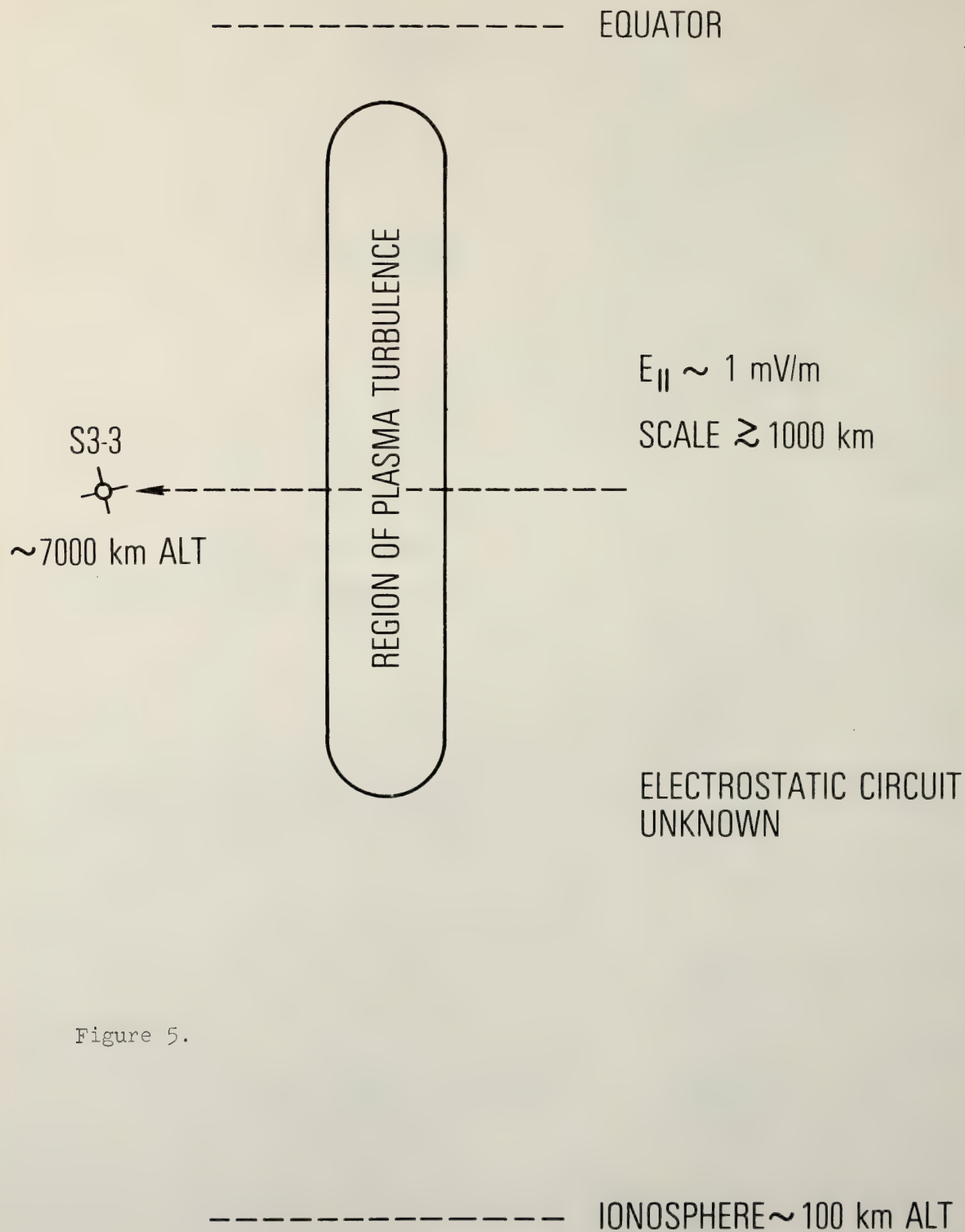


Figure 5.

The fourth viable mechanism for maintaining a magnetic field-aligned electric potential drop is that due to the magnetic mirroring effects of differential pitch-angle anisotropy between ions and electrons (Alfvén and Fälthammar, 1963). Unlike the other mechanisms considered above, this mechanism depends on the magnetic field structure being suitable for mirroring of the energetic plasma injected into the auroral region, for if the equatorial pitch-angle distributions of such ions and electrons are different their "average" mirroring locations will be different, thus setting up a charge separation electrostatic field. A number of authors have considered such a mechanism for the case of auroral plasma (e.g., Lemaire and Scherer, 1974; Whipple, 1977; Lennartsson, 1977; Chiu and Schulz, 1978), assuming strict charge neutrality. This is one model where a careful consideration of the contribution of cold ionospheric electrons is absolutely essential. One-dimensional quasi-neutral calculations (Chiu and Schulz, 1978) indicated that ionospheric plasma is crucial in the magnetic mirroring mechanism not only in giving a proper account of electron distributions, as in the phenomenological model of Evans (1974), but also in partially short-circuiting the very large potential drops expected from consideration of magnetospheric plasma alone (Alfvén and Fälthammar, 1963). In any case, the parallel scale length of this mechanism is essentially the field line distance between the ionosphere and the magnetospheric equator, i.e., the region in which the plasma mirrors, yielding parallel electric fields well below 1 mV/m (see Figure 6).

There is yet another possible source of auroral electric fields that accelerate ions and electrons in opposite directions along the earth's magnetic field. This last possible source is the magnetospheric convection electric field. The convection electric field is perpendicular to the magnetic field at high altitudes, but its meridional ( $r, \theta$ ) component has a theoretical discontinuity at or near the boundary between closed and open magnetic field lines (see Figure 7, which shows the amplitudes of the diurnal variation of  $E$  at ionospheric altitudes). Ionospheric resistivity would partially connect electrostatic equipotentials across the discontinuity, but at too low an altitude to account properly for the observed component of  $E$  parallel to  $B$ . However, the "kinematical resistivity" associated with magnetic-mirror forces on a hot plasma may increase the altitude at which the parallel (to  $B$ ) component of  $E$  would appear. The details of this latter effect, which (if it occurs) would produce the desired distribution of  $E \cdot B$  with altitude, remain to be worked out. However, the effect would be such as to produce an upward electric field in the PM sector (maximal at dusk) and a downward electric field in the AM sector (maximal at dawn) of the auroral oval. This expectation is in good agreement with the diurnal distribution of upgoing ion beams observed by Ghielmetti et al. (1978).

It is evident that these physical mechanisms do not exist entirely independently of one another. For example, if the restrictive assumption of strict charge neutrality is removed in the magnetic-mirror model, one has an oblique electrostatic shock in a mirroring field. To the extent that no dissipative mechanisms such as wave-particle turbulence are included in such a "shock", the resulting electric-field structure is better described as a solution of Poisson's equation. From another point of view the oblique shock can be described as some sort of zero-frequency electrostatic ion-

# MAGNETIC MIRRORING

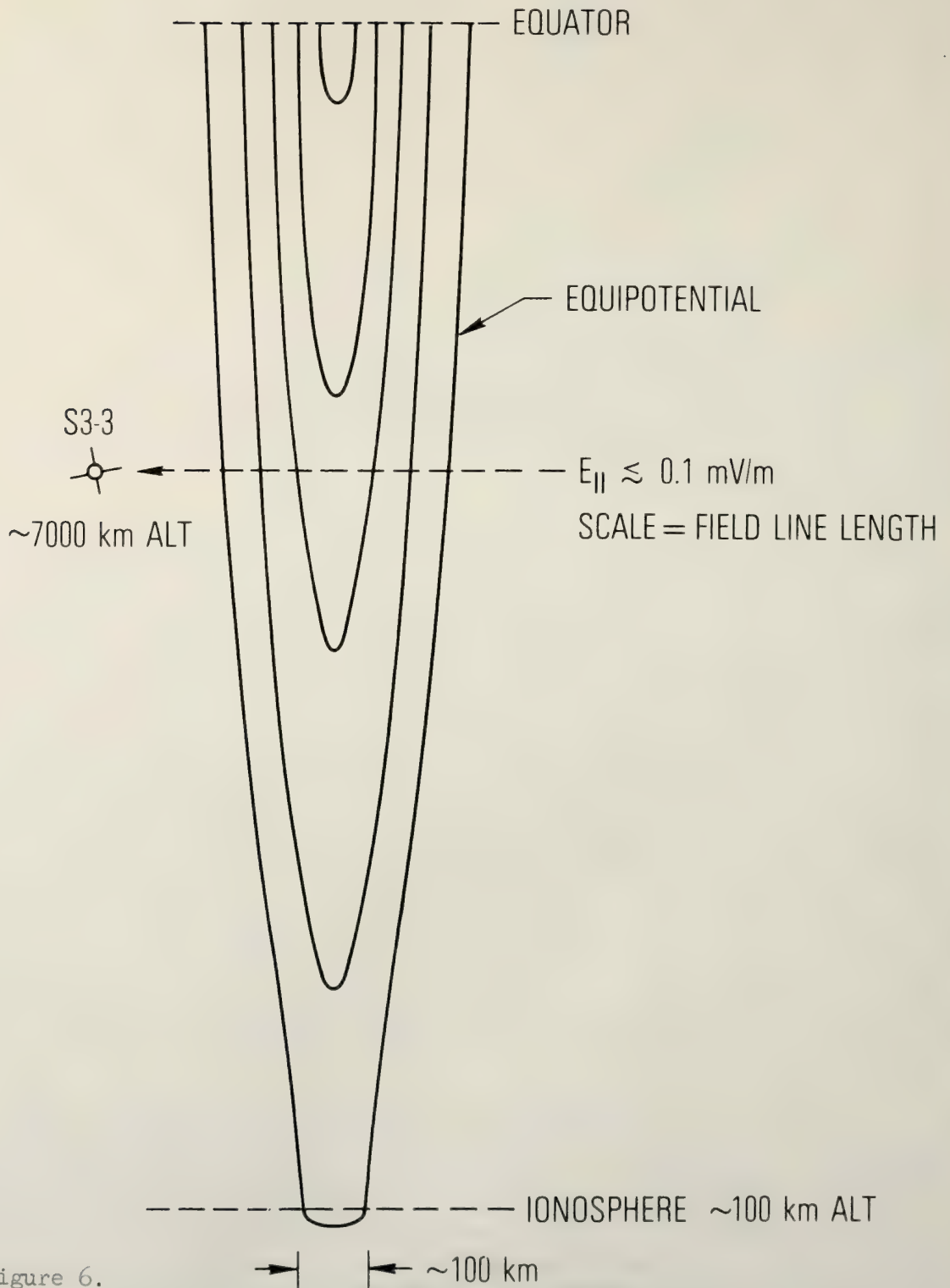


Figure 6.



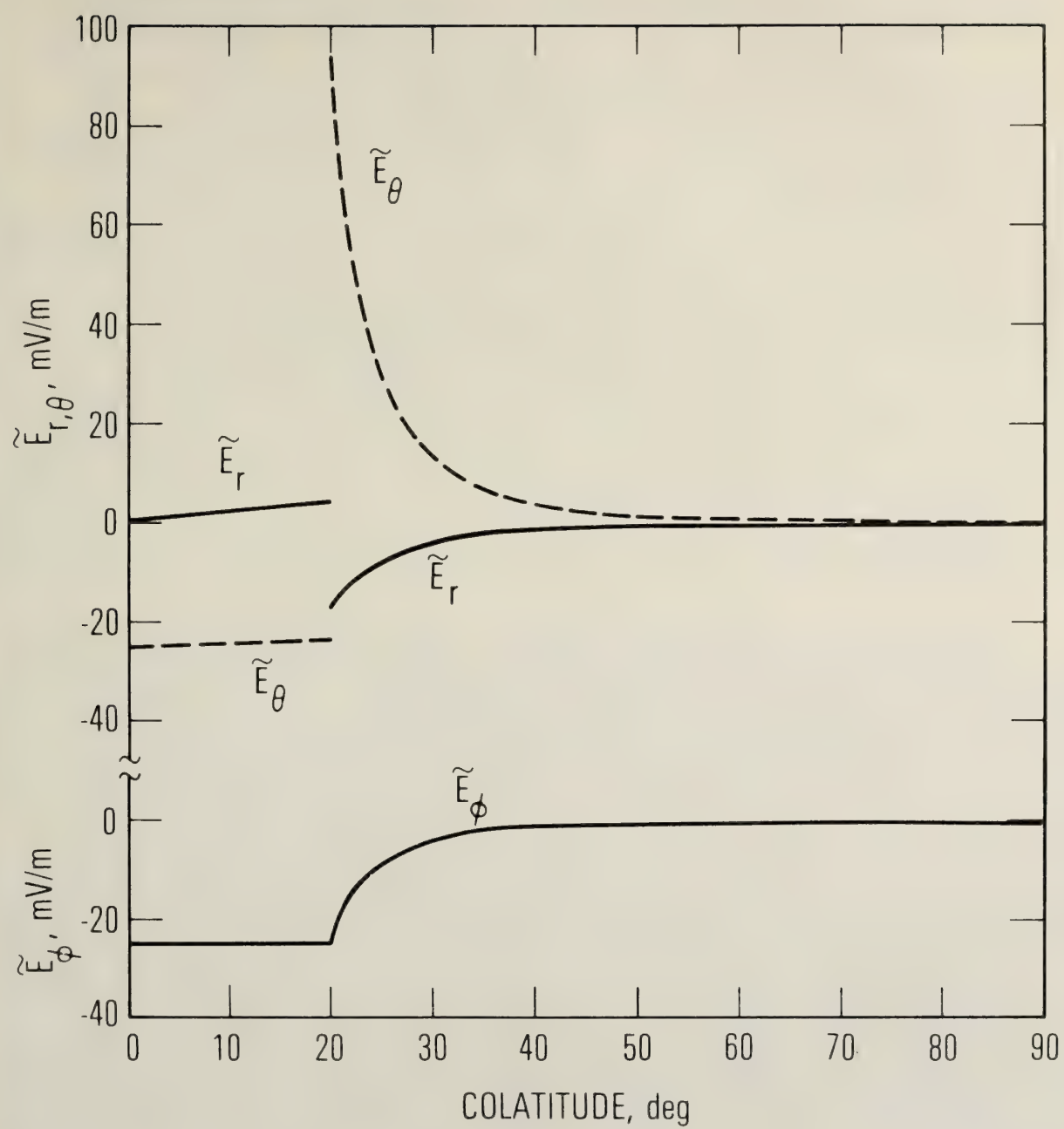


Figure 7

cyclotron (EIC) mode. There surely is a great deal of EIC turbulence connected with auroras, and the physical distinction between the oblique shocks of Swift and this turbulence is at best imprecise. Yet the merging of wave turbulence and shocks can lead to substantial parallel electric potential drops in the complete absence of anomalous resistivity. (The reader need not be reminded that turbulence is not synonymous with anomalous resistivity; in fact, it is quite difficult to make anomalous resistivity out of even the most turbulent waves.)

In a recent work, Chiu and Cornwall (1978) have considered Poisson's equation in dipolar magnetic geometry, coupled with ionospheric physics. In such a model the parallel potential drop is intimately coupled to the perpendicular electrostatic field structure as indicated in Figure 1. Further, the scale length of the perpendicular electrostatic field structure is related not only to the field-aligned current to the ionosphere but also to the ionospheric Pedersen conductivity. Thus, ionosphere-magnetosphere coupling is a crucial ingredient determining the geometric structure as well as the energetics of the quiet auroral arc in such a model. At present, no satisfactory solution of such a model has yet been obtained in the return-current region, although an approximate solution in the central electron beam region has been obtained. A schematic illustration of this model is given in Figure 6.

Because the various mechanisms are not necessarily mutually exclusive, it is difficult to rule out any particular mechanism by observations; however, the parallel scale length, or equivalently, the peak magnitude of the parallel electric field, may be used to distinguish some models from others. Since the current-driven double layer is distinguished by a very short parallel scale length, one may ask if the parallel (to B) electrostatic-field observations of S3-3 would be able to distinguish the double layer from other mechanisms. Mozer et al. (1977) reported very large parallel electrostatic fields ( $\geq 100$  mV/m) in the presence of  $> 100$  mV/m perpendicular electrostatic fields. These have been identified as  $\sim 800$  mV/m parallel electrostatic fields of double layers (Shawhan et al., 1978). Hudson and Mozer (1978) were cautious in making such an identification because "the angular resolution of the instrument may alias the parallel electric field measurement in the presence of strong perpendicular electric fields greater than 100 mV/m." Particle measurements on S3-3 cannot resolve the question of parallel scale length either, although they do put constraints on double-layer models such as the necessity of multiple formation pointed out earlier. Thus, for the time being, no mechanism discussed above can be ruled out, but it must be said that double layers are unlikely both theoretically and experimentally. Clearly, further theoretical development of these models is needed to help the process of experimental elimination of unsuitable candidates. Identification of the auroral mechanism is especially important for ionospheric-magnetospheric coupling studies because the ionosphere plays roles of varying importance in various mechanisms. It would be very hard to believe that the ionosphere plays no active role at all in the dynamics of the aurora. In such an eventuality, ionospheric currents would be entirely decoupled from the magnetospheric currents, and our growing understanding of the relationship between auroral dissipation of currents and energy input into the magnetosphere would be lost. On the contrary, it seems that there is a genuinely strong

coupling between magnetospheric, auroral, and ionospheric phenomena.

#### IV. PROSPECTS

As we have seen in previous sections, the S3-3 measurements not only confirm the suspected existence of kilovolt electrostatic potential drops along auroral field lines but also clarify the relationship between inverted-V structures, electron and ion beams, and electrostatic fields in the aurora. However, because of aliasing problems, the parallel electric field measurements on board are unable to determine the parallel scale length of the electrostatic potential drops with confidence. We would have to depend on future experiments to settle this crucial question. If parallel electrostatic fields are as large as 800 mV/m, as one experiment suggests, then there is no question that some sort of double layer or electrostatic shock with small obliquity is involved. If on the other hand, the parallel electrostatic field turns out to be  $\leq 1$  mV/m, as many experiments suggest, then anomalous resistivity, oblique shocks, and magnetic mirroring are all candidates. Undoubtedly these three mechanisms go hand in hand, so it is not a question of choosing only one of them.

Theorists are not yet ready to pronounce judgment in favor of one or another model, thus leaving open one vital question: Do parallel electric fields isolate the magnetosphere from the ionosphere? We (in agreement with traditional views) think not, but we know of no definitive experimental evidence which shows how magnetospheric and ionospheric current paths are closed. It may be that radar and other ground-based studies combined with satellites such as S3-3 can provide this evidence.

Another important issue is the mapping of perpendicular electric fields along field lines from the magnetosphere to the ionosphere. For most models other than the double layer, the mapping modifications induced by parallel electric fields may not be terribly significant, but there may be almost complete decoupling in double-layer models, so the mapping question will have to be completely reexamined. Presently, calculations of ionospheric currents, and their concomitant heating of the thermosphere (Straus and Schulz, 1976), depend on a variety of electrostatic models which depend on direct mapping of convection electric fields (e.g., Volland, 1975). If double layers exist over regions as extensive as inverted-V structures (to explain the electron beams), then the question of how convection electric fields map through a double layer must be addressed. Indeed, the role of the observed paired perpendicular auroral electrostatic fields (Figure 1) has not been considered in double layer theory.

Proponents of active ionospheric coupling with the magnetosphere will surely note that the observation of oxygen ion beams on S3-3 is evidence that the ionosphere may be a major source of charged particles for the auroral magnetosphere. Very recent isotopic-ratio observations of ring-current ions in the magnetosphere indicate that the ionosphere (via the aurora) may be a major source for the ring current as well (e.g., Young et al., 1977). We suggest that the next major advance in ionosphere-



magnetosphere coupling may be the understanding, both observational and theoretical, of the ultimate fate of these ion beams. Since downward moving ion beams are not observed at high altitudes, and since conical beams are primarily an ionic phenomena, the question of what happens to ion beams before they reach the equator appears to be interesting indeed.

#### ACKNOWLEDGMENT

This work is partially supported by NASA under Grant NASW-3120 and partially by The Aerospace Corporation Sponsored Research Program.

#### REFERENCES

- Alfvén, H., and C. -G. Fälthammar (1963): Cosmical Electrodynamics, pp. 163-167, Clarendon Press, Oxford.
- Armstrong, J. C., and A. J. Zmuda (1973): Triaxial magnetic measurements of field-aligned currents at 800 km in the auroral region: Initial results, J. Geophys. Res., 78, 6802.
- Block, L. P. (1975): Double layers, in Physics of the Hot Plasma in the Magnetosphere, pp. 229-249, edited by B. Hultqvist and L. Stenflo, Plenum Press, N. Y.
- Cauffman, D. P., and D. A. Gurnett (1972): Satellite measurements of high latitude convection electric fields, Space Sci. Rev., 13, 369.
- Chiu, Y. T., and J. M. Cornwall (1978): Electrostatic model of a quiet auroral arc, submitted to J. Geophys. Res.
- Chiu, Y. T., and M. Schulz (1978): Self-consistent particle and parallel electrostatic field distributions in the magnetospheric-ionospheric auroral region, J. Geophys. Res., 83, 629.
- Cloutier, P. A. (1971): Ionospheric effects of Birkeland currents, Rev. Geophys. Space Phys., 9, 987.
- Croley, D. R., Jr., P. F. Mizera, and J. F. Fennell (1978): Signature of a parallel electric field in ion and electron distributions in velocity space, J. Geophys. Res., 83, 2701.
- Evans, D. S. (1974): Precipitating electron fluxes formed by a magnetic field-aligned potential difference, J. Geophys. Res., 79, 2853.
- Evans, D. S. (1975): Evidence for the low altitude acceleration of auroral particles, in Physics of the Hot Plasma in the Magnetosphere, pp. 319-340, edited by B. Hultqvist and L. Stenflo, Plenum Press, N. Y.

- Frank, L. A., and K. L. Ackerson (1971): Observation of charged particle precipitation into the auroral zone, J. Geophys. Res., 76, 3612.
- Ghielmetti, A. G., R. G. Johnson, and E. G. Shelley (1978): The latitudinal, diurnal and altitudinal distributions of upward flowing energetic ions of ionospheric origin, Geophys. Res. Letters, 5, 59.
- Heppner, J. P. (1972): Electric field variations during substorms: Ogo 6 measurements, Planet. Space Sci., 20, 1475.
- Hudson, M. K., and F. S. Mozer (1978): Electrostatic shocks, double layers, and anomalous resistivity in the magnetosphere, Geophys. Res. Letters, 5, 131.
- Hudson, M. K., R. L. Lysak, and F. S. Mozer (1978): Magnetic field-aligned potential drops due to electrostatic ion cyclotron turbulence, Geophys. Res. Letters, 5, 143.
- Kan, J. R. (1975): Energization of auroral electrons by electrostatic shock waves, J. Geophys. Res., 80, 2089.
- Kindel, J. M., and C. F. Kennel (1971): Topside current instabilities, J. Geophys. Res., 76, 3055.
- Kintner, P. M., M. C. Kelley, and F. S. Mozer (1978): Electrostatic hydrogen cyclotron waves near one earth radius altitude in the polar magnetosphere, Geophys. Res. Letters, 5, 139.
- Lemaire, J., and M. Scherer (1974): Ionosphere-plasmasheet field-aligned currents and parallel electric fields, Planet. Space Sci., 22, 1485.
- Lennartsson, W. (1977): On high-latitude convection field inhomogeneities, parallel electric fields and inverted-V precipitation events, Planet. Space Sci., 25, 89.
- Mizera, P. F., and J. F. Fennell (1977): Signatures of electric fields from high and low altitude particle distributions, Geophys. Res. Letters, 4, 311.
- Mizera, P. F., D. R. Croley, Jr., and J. F. Fennell (1976): Electron pitch-angle distributions in an inverted-'V' structure, Geophys. Res. Letters, 3, 149.
- Montgomery, D. C., and G. Joyce (1969): Shock-like solutions of the electrostatic Vlasov equation, Plasma Phys., 3, 1.
- Mozer, F. S., and R. H. Manka (1971): Magnetospheric electric field properties deduced from simultaneous balloon flights, J. Geophys. Res., 76, 1697.

- Mozer, F. S., C. W. Carlson, M. K. Hudson, R. B. Torbert, B. Paraday, J. Yatteau, and M. C. Kelley (1977): Observations of paired electrostatic shocks in the polar magnetosphere, Phys. Rev. Letters, 38, 292.
- Papadopoulos, K. (1977): A review of anomalous resistivity for the ionosphere, Rev. Geophys. Space Phys., 15, 113.
- Sharp, R. D., R. G. Johnson, E. G. Shelley, and K. K. Harris (1974): Energetic  $O^+$  ions in the magnetosphere, J. Geophys. Res., 79, 1844.
- Shawhan, S. D., C. G. Fälthammar, and L. P. Block (1978): On the nature of large auroral zone electric fields at  $1-R_E$  altitude, J. Geophys. Res., 83, 1049.
- Shelley, E. G., R. D. Sharp, and R. G. Johnson (1976): Satellite observations of an ionospheric acceleration mechanism, Geophys. Res. Letters, 6, 54.
- Shelley, E. G., R. G. Johnson, and R. D. Sharp (1972): Satellite observation of energetic heavy ions during a geomagnetic storm, J. Geophys. Res., 77, 6104.
- Straus, J. M., and M. Schulz (1976): Magnetosphere convection and upper atmospheric dynamics, J. Geophys. Res., 81, 5822.
- Swift, D. W. (1975): On the formation of auroral arcs and acceleration of auroral electrons, J. Geophys. Res., 80, 2096.
- Swift, D. W. (1976): An equipotential model for auroral arcs, 2, numerical solutions, J. Geophys. Res., 81, 3935.
- Volland, H. (1975): Models of global electric fields within the magnetosphere, Ann. Geophys., 31, 154.
- Vondrak, R. R., H. R. Anderson, and R. J. Spiger (1971): Rocket-based measurements of particle fluxes and currents in an auroral arc, J. Geophys. Res., 76, 7701.
- Whipple, E. C., Jr. (1977): The signature of parallel electric fields in a collisionless plasma, J. Geophys. Res., 82, 1525.
- Young, D. T., J. Geiss, H. Balsinger, P. Eberhardt, A. Chielmetti, and H. Rosenbauer (1977): Discovery of  $He_2^+$  and  $O_2^+$  ions of terrestrial origin in the outer magnetosphere, Geophys. Res. Letters, 4, 561.



## B. HIGH-LATITUDE E- AND F-REGION IONOSPHERIC PREDICTIONS

A WORKING GROUP REPORT prepared by R. Hunsucker, Chairman, and Members: R. Allen, P. Argo, R. Babcock, P. Bakshi, D. Lund, S. Matsushita, G. Smith, A. Shirochkov, and G. Wortham.

### 1. INTRODUCTION

WG C2 is concerned about the E and F layers at latitudes greater than  $\Lambda \approx 60^\circ$ . Specifically, we address the pertinent physical processes, the morphology, the existence and adequacy of models, and the reliability of predictions for the high-latitude E and F layers at the present time and in the foreseeable future.

In order to analyze this problem we must:

1. As completely as possible, describe the temporal and spatial behavior of the high-latitude ionosphere--specifically, the disturbance aspects: sporadic E, spread F, extent of trough, location of cusp, etc.

2. Determine how well we can predict the real-time location of the auroral oval.

3. Consider carefully the ionospheric reflection points for each HF propagation path to be analyzed in relation to the auroral oval or disturbance region.

4. Consider how best to use the next generation instrumentation (satellite, NOAA HF radar, Polar Cusp Incoherent Scatter Radar, EISCAT, etc.) data to improve E- and F-region predictions.

5. Match, as closely as possible, user needs to realistic prediction capabilities.

6. Fully utilize data produced by new observatories on polar cap/cusp ionospheric regions (Cape Parry, Sondrestromford, Spitzbergen).

This working group considers that our starting points are in the specific needs of our users. Systems users now need predictions for both transionospheric and reflected propagation effects in the high latitude ionosphere. Navigation systems need propagation time delay, radars need range and angle corrections. Communication systems, both broadcast and point-to-point need more than availability; they also need predictions for multimode, dispersion, angle of arrival, received noise to signal, etc. In essence, users of advanced systems need and demand predictions tailored to their systems and interpreted directly in their own system parameters.

At the system design stage, the potential user needs long-term prediction of the expected worst conditions, as well as climatological medians. In most proposed systems, the basic data on pertinent ionospheric parameters do not exist in a synoptic form; therefore, long-term prediction for advanced system design cannot be made with reliability.

Consider two advanced systems at the concept stage. The first is a

short haul (200-700 km) extended HF system to be designed to provide high density digital data, using chip interleaving and error correction. The design might depend critically on a scattering cross section function for auroral/polar cap sporadic E (Es) and the expected fading distribution. The second is a navigation system for ships, aircraft, and remote ground exploration, using coded transmissions on two channels from the same satellite. The accuracy of the position determination might depend critically on the spatial rate of change of the ionosphere when the measurements are made sequentially to the moving satellite. In addition, F-region fading could prevent locking or relocking on the coherent code.

For systems already fielded, the users require long-term predictions for system management, e.g., frequency allocation. Some users are benefited by predictions for monthly median effects, for instance using predicted time delay or ionospheric range error to reduce their monthly rms variability by a factor of 3 to 4. Most users are receptive to alerts for expected extremes in the ionospheric response of their systems, such as blackout, increased fading, etc., even if they cannot do anything about it. Some advance systems may require slow-acting adaptive systems with an external sensor of current ionospheric conditions so that they can remove most of their ionospheric effects in near real time.

## 2. PHYSICAL PROCESSES IN THE HIGH-LATITUDE E AND F REGIONS

### 2.1 The Present State of Knowledge of Physical Processes in the High Latitude E and F Regions

In the mid- and low-latitude E and F regions, the primary source of ionization is solar EUV and UV. The day-to-day variations in the ionosphere essentially reflect changes in the neutral atmosphere density and winds. Electric fields occasionally become an important modulating force during disturbed periods. At high latitudes, the magnetosphere-ionosphere interaction is the dominant dynamic force. There are two primary processes. The precipitation of energetic particles is an important source of ionization, particularly in the E region. The strong electric fields which arise from the interaction of the interplanetary magnetic field (IMF) and the geomagnetic field drive a two-celled ion convection pattern with anti-sunward ion flow across the pole and a return flow along the auroral oval. At E-region heights, the ion convection is restrained by collision with neutrals. The result is a net current flow known as the auroral electrojet which creates Joule heating of the neutral atmosphere. Changes in the neutral density and winds associated with this heat act as feedback mechanisms that change the density and structure of the high-latitude ionosphere.

Each of the main features of the high latitude E and F regions is related to particular aspects of the magnetospheric-ionospheric interactions.

The auroral oval is an active region of particle precipitation and electric currents. The most obvious characteristic of the oval is the visible light emissions that are used to demarcate its boundaries. The equatorward edge of the oval marks the boundary between the magnetic field lines that are essentially dipolar, known as the plasmasphere, and the field lines that are interacting with the IMF and are being dragged into the magnetotail.



Most of the particle precipitation which creates the visible aurora occurs in the nighttime sector. The auroral region is also thickest in latitude in the nighttime sector. In the day sector, the aurora becomes weaker, with less latitudinal extent; during very quiet times it is often not seen.

The aurora seen near the equatorward edge of the oval and most of the daytime aurora is known as diffuse aurora. It is caused by the precipitation of particles with a relatively broad spectrum of energies (e.g., electrons of 100 eV to 100 keV). Most of the energy results in strong ionization in the E region, creating a broad shelf in the electron density profile. But the lower energy particles also enhance the F-region density. Poleward of the diffuse aurora is the region of discrete auroral arcs. These are associated with monoenergetic particle beams which create a "bubbling" effect in the E region at the end of the field lines along which they are traveling. The result is a thin layer of very enhanced electron density in the E region.

During disturbed periods the auroral oval expands equatorward, the intensity of the diffuse aurora and the number and strength of the discrete arcs increases. The appearance of the westward traveling surge and the breakup of the discrete arcs results from a complex pattern of particle precipitation and broad regions of enhanced E-region densities. There is also an increase in the intensity of the daytime aurora.

Another region of particle precipitation is the dayside cusp or cleft. The cusp is typically 2 to 4 degrees wide in longitude and located at about  $\Lambda = 78-80^\circ$  near local noon. The dynamics of the cusp are not as clearly known as for the nighttime auroral zone. It appears that the precipitation seen in the cusp region is related more to the direct interaction of the magnetosphere and the IMF. The precipitation tends to be less energetic and to vary less during disturbed periods. Enhancements in electron density at all heights have been observed.

A third feature of the E and F region associated with the high latitude ionosphere is the main electron density trough. This is a region of markedly decreasing electron density as one moves from mid to high latitudes in the nighttime ionosphere. At the poleward boundary of the trough there is a sharp increase in density due to the ionizing effect of the auroral precipitation. The main trough is primarily a winter nighttime phenomena. It appears to be caused by increased recombination due to the shorter daytime ionization period and the extended period the ions are kept in the nighttime by the ion convection pattern. At higher altitudes there is also a light ion ( $H^+$ ) trough which has been observed by satellites. This trough appears to be related to the loss of ions of ionospheric origin into the magnetic tail. The light ion trough and the main trough are not always coincident.

The last major feature of the high latitude ionosphere is the polar cap. Compared to the highly dynamic auroral zone, the polar cap is rather uncomplicated. However, the available observations of the polar cap region are relatively few, so our understanding of this region is also limited. Discrete auroral arcs parallel to the ion convection pattern have been observed by satellite and the polar cap is affected by energetic solar particles that cause radio blackouts, although this is essentially a D-region effect.

## 2.2 Where We Need To Go

The processes occurring in the auroral oval are the most dynamic and exciting and are clearly the best understood. With the advent of satellite ob-



servations, the physical processes that drive the various auroral phenomena have been examined fairly extensively. There is a good understanding of these processes and their relation to the various features of the auroral zone appearing in morphological models. Within the auroral zone, the variations in the F region are probably the least known phenomena, particularly during the disturbed periods. The strong E region created by particle precipitation has limited the ability to observe the F region using ground-based observations. The expanded use of transionospheric propagation in high-latitude regions makes knowledge of F-region variations of increasing importance.

The other main features of the high latitude ionosphere have only recently become the focus of concerted research efforts. The dynamics of the main ionospheric trough and the light ion trough are not well understood. There is no clear consensus on how and why they are formed, and how, or if, they are related and interact. The number of observations of the cusp region is still small enough that there is no clear picture of exactly what the morphology of the energies of the precipitating particles is and how precipitation varies with geomagnetic activity. The dynamics of the cusp as it moves with variations in the IMF is not well understood. The polar cap is the least active of the high-latitude regions, but it is also the least observed and the least understood.

Another important factor in the dynamics of the high-latitude ionosphere that is not well understood, and often ignored by ionospheric physicists, is the strong interaction between the ionosphere and the neutral atmosphere. Changes in the neutral density and winds that are driven by ionospheric processes not only are a feedback into the high latitude ionospheric dynamics, but the energetics of auroral processes are an important driving force of the global thermospheric circulation and will indirectly affect the ionosphere at lower latitudes.

### 3. MORPHOLOGY

#### 3.1 Where We Are

From the point of view of the geophysicist, a lot still remains to be done before we have a clear understanding of the morphological features of the high-latitude E and F regions. The best determined feature of this region is the position of the auroral oval, determined by visual observations, all-sky camera data, and, more recently, satellite photographs. Almost as well defined are three auroral absorption regions, one just equatorward of the diffuse aurora in the midnight sector, another in the pre-noon sector, and a third in the early morning sector during disturbed periods. All have been studied extensively, principally by an international network of riometers but recently by in-situ particle measurements. Both the auroral oval and the auroral absorption regions are well represented by statistical (computer version) models.

Our knowledge of the main ionospheric trough is in much worse condition. Although it has been studied since the previous solar minimum and has been demonstrated in archived ionosonde data, little is known of its variations

with solar cycle, its disturbance behavior, the positional variations of the plasmopause, and the variation of ionospheric parameters over this feature. Lack of consistent synoptic data has prevented development of more than first-order statistical models of an F-region trough.

Very much more must be done in studying the localized features of the E and F2 regions, particularly for the polar region of the Southern Hemisphere. The behavior of the F2 region during magnetospheric substorms is so dramatic that we now know it comparatively well on the night side; however, our studies are very incomplete on the day side. For very strong ionospheric disturbances, e.g., PCA or strong aurora, our data from ground-based ionosondes is almost always blacked out; hence, we know even less about the F region during those periods. Recent data from instruments, such as incoherent scatter and in-situ probes, are beginning to identify and delineate some of the critical parameters of the high-latitude E- and F-region features, but there is not yet sufficient synoptic data to build either descriptive or predictive models.

From the point of view of the system engineer, the morphology should address the effects of the high-latitude ionosphere on his particular system in terms of his system's parameters. He almost always would like these derived from a comprehensive synoptic data base identical to his system. It is very doubtful that this exists, even for standard narrow band HF communication. Even parameters of critical design to slightly modified systems, such as joint availability of various E and F modes, exist only in restricted time-space segments. In terms of those features that are of critical importance to advanced systems, like spread-F fading statistics for the cusp region or the poleward edge of the trough, the scattering cross section of sporadic E in the midnight sector and slant E in the polar cap, not very much high quality data exists.

As a result morphological descriptions of features important to the system designer and user do not exist except in very simplistic qualitative form.

### 3.2 What Is Needed

There is a critical need for multinational, multidiscipline experiments. Since we are just beginning to understand the physical processes by which solar energy enters and cascades through the upper atmosphere/ionosphere, simultaneous measurements are mandatory. In particular, cooperative attention should be given to the "hot points" at the cusp and at the main trough in both the north and south. Such research must be done on a new level. All physical sources of maintaining and varying the high latitude must be found and studied. Therefore, in-situ measurements must be made of the parameters of the solar blast wave, followed by particle, optical radio magnetic and electric field measurements to determine the energy deposition at the cusp, morningside and midnight sectors. New generation ground-based radio instruments, such as incoherent scatter and advanced (HF radars) and modern satellite sensor of in-situ and integrated ionospheric effects must determine the dynamical behavior of the E and F regions, particularly during very large solar energy storms when archive data are totally inadequate. For the benefit of existing and proposed system users, simultaneous radio path propagation conditions should be monitored by several classes of instruments, such as oblique sounders to determine mode availability, MUF and



LUF and fixed frequencies to determine the status of dispersion, fading, oblique absorption and scattering loss through all phases of geomagnetic storms.

If an adequate morphology of the ionosphere is to be developed, such coordinated experiments must be conducted on the day and night sides and in both hemispheres.

#### 4. MODELS

Models are required for long-term prediction of propagation for ionospherically supported systems (VLF, HF, and VHF scatter) and for satellite systems. This working group has focused on predictions for HF systems, since the D region (which determines VLF propagation) and F-region irregularities (which cause satellite scintillations) are being addressed by other groups. However, the results should be combined.

Among the high-latitude ionospheric features that should be modeled for HF predictions are the following:

- a. Location and extent of auroral oval as a function of magnetic activity.
- b. E- and F-region parameters in auroral oval, midlatitude trough, daytime cusp, and polar cap.
- c. F-region parameters, important for ducting.
- d. Profile of  $N_e$  versus height between E and F2 regions, of importance for prediction of ducted propagation.
- e. Auroral absorption, PCA absorption (D-region effects).
- f. E- and F-region irregularities that cause diffuse multipath (spread F), rapid flutter fading, and off-great-circle propagation.

In the auroral region, models must be detailed enough for raytracing. That is, the shape of the profile and the steep gradients must be represented in three dimensions. Predictions of MUF/LUF using virtual geometry (a priori mode definition and location of reflection points) are meaningless because strong horizontal gradients cause large changes in mode geometry and therefore in D-region crossing points and absorption estimates.

##### 4.1 Existence of Models ("Present")

Several models are available for the six features listed above. The ITS-78 (also Miller-Gibbs) model is usually taken as the basic model in the United States and modified either by superposition of more detailed features or by revision of the coefficients themselves. The ITS model is relatively poor in the auroral region. Data accumulated with ground-based and airborne ionosondes by AFGL and RADC (formerly AFCRL) have been used to develop an improved model. This model has recently been further modified by SRI using incoherent-scatter radar (ISR) data from the Chatanika radar near College, Alaska. The SRI model retains the foF2 and hmF2 values in the RADC 1976 Polar Model and incorporates an auroral E layer (essentially equivalent to blanketing  $E_s$ ) and valley from ISR data, as well as an absorption model that varies with magnetic activity.

The USSR has developed statistical models (see, for example, Kovalevskaya



and Zhulina, 1979; Troshichev et al., 1979; Avdiushin et al., 1979; and Kerblay and Nosnova, 1979). Differences have been noted between their predictions and those of the United States in the case of auroral absorption and foF2 in the midlatitude trough during solar maximum. During magnetic disturbances, three distinct regions of auroral absorption have been identified by Soviet scientists, whereas the recent SRI absorption model displays two time/latitude regions of enhanced absorption. Gradients of foF2 in the trough at solar maximum were found to be large by the Soviet group, but relatively small by RADC [the ARCON model by Miller and Gibbs (1975)]. Further model verification is clearly needed.

The three-dimensional models developed by RADC and SRI are suitable for raytracing. In the original RADC model, the Ne(h) profile between layers must be assumed; the SRI model attempts to match profile shapes actually observed with the ISR radar. Off-great-circle propagation by refraction can be predicted by three-dimensional raytracing. Separate computer codes are available for off-great-circle propagation by scattering from field-aligned irregularities; however, the usable frequencies and signal strength are not predicted.

#### 4.2 Adequacy of Models ("Future")

The capabilities and limitations of the U.S. models are summarized in Table 1.

### 5. PREDICTIONS

#### 5.1 Present State of Predictions

The ability to predict the propagation of HF radio waves along auroral and polar paths requires a model of the spatial distribution of electron density in the high-latitude ionosphere; one must then raytrace through this model to calculate the raypath connecting the end points. Because of the very definite spatial structure in the auroral and polar cap regions, the specific propagation predictions may be very sensitive to the positioning of the structures. It is difficult to determine whether washed out "strongly average morphologies" will give more useful results than misplaced features--averaged test results do compare better with averaged morphologies. Many system designers, however, are likely to be concerned with "worst-case" occurrence estimates, which are more accessible through even misplaced structures.

Several different types of systems are dependent upon propagation through the high-latitude ionosphere. Table 2 specifies some of these systems and the type of information needed for the systems to operate optimally.

The present prediction capability is statistical, using averaged synoptic data to define morphologies that ignore small scale (<100 km) structure and transient phenomena (<3 hr). Therefore critical features such as auroral arcs and substorms are completely neglected. The models do contain auroral E, some auroral E<sub>s</sub>, auroral absorption, and the F1 and F2 regions.

HF systems designers and communicators are concerned with the propagation facets mentioned in Table 2. The availability, multimode structure,

Table 1. U.S. Ionospheric Models

Feature	Prediction	Adequacy	Techniques for Potential Improvement
1. Auroral oval	1. Feldstein oval for E region (Extent differs for D and F regions)	1. Approximately equivalent to optical predictions 2. Good on 3-hour average	1. DMSP optical for real time (already done) 2. Particle counters (satellite) 3. Chatanika, EISCAT ISR 4. Backscatter radar (Bates' work) 5. Synoptic analysis of bottom-side and topside ionograms
2. E, F2 parameters	ITS coefficients, with RADC modifications  SRI foEs model (Chatanika data)	1. foF2 good except during storms 2. hmF2 contains anomalies 3. Good within limits of data base used if oval location predicted correctly	1. Analysis of more ISR data 2. Analysis of existing ionograms 3. Comparison of ISR and ionogram data 4. Photometers 5. Particle counters
3. F1 layer	1. X and SSN analytic formula	1. Fair	1. Petrie and Stevens 1965 model 2. ISR data analysis 3. Ionogram analysis, synoptic
4. Ne(h) profile	1. Analytic, assumed 2. SRI Chatanika model	1. Poor 2. Good within limits of data base	1. ISR data analysis and verification 2. Rockets (3) topside sounding
5. Auroral	1. Foppiano A. A. model  2. SRI modification of Foppiano model	1. No dependence on magnetic activity 2. Has Kp dependence but has not been verified against independent data	1. Riometer data available from several chains should be used for verification 2. A-1 absorption measurements 3. Oblique path HF signal strength
6. Irregularities (spread F)	None in HF ray trace codes. Field-aligned scatter geometry codes available	General morphology known but not applied in prediction (e.g., Davies, 1972; and Singleton, 1979.	1. Signal strength measurements coordinated with VI observations 2. Satellite scintillations and spread-F ionogram correlation needed, both simultaneous and statistical.



Table 2. Systems and Prediction Needs

System	Prediction Needs
HF Comm	availability, dispersion, multi mode, fading, noise, absorption, scattering cross section, ducting, non-great-circle propagation
VHF/UHF navigation and communication	Fading (amplitude, phase), $\Delta\text{TEC}/\Delta\text{lat}$ , angular deviation
remote sensing (HF)	Refraction, diffraction
VLF navigation	PCA phase advances; auroral D-region effects

ducting, absorption, noise, and non-great-circle propagation can be dealt with in some averaged, statistical sense (see, for example, Hatfield, 1979) that is useful to the communications system designer. Since the auroral/polar cap models do incorporate varying magnetic activity conditions, the systems designer can test his system under various scenarios.

The accuracy of the predictions available through the use of these polar and auroral models has not been evaluated--it is expected that on the average they will be significantly worse than the midlatitude ITS-78 predictions (Barghausen et al., 1969). This is because of the extreme sensitivity of the high-latitude ionosphere (with its many discrete features) to geomagnetic activity. Therefore, it is not likely that present day prediction capabilities would be very useful to the communication planner (e.g., frequency allocation) on a day-to-day or even month-to-month basis.

Remote sensing systems (HF radars) are also very dependent upon propagation through the polar/auroral ionosphere. The refraction and diffraction of signals by the discrete features will strongly affect the operation of these systems. In this sense, the usefulness of predictions for these systems is on approximately the same level as for HF communications.

Because the very high frequency transionospheric radio signals used in VHF navigation (e.g., GPS) are only slightly refracted (or time delayed) by the polar/auroral E region, the main effects are F-region refractions and auroral scintillations. The dominant F-region effect is steep spatial gradients; therefore, under the highly disturbed oval, knowledge of its location is of primary importance. This is probably best measured in real time (as the southernmost edge of the auroral oval). The scintillation morphology is presently a statistical model (this area is covered in detail by working group E2 of the ISTP Workshop).

VLF navigation (e.g., Omega) is very dependent upon the condition of the polar D region, and is only marginally affected by the E and F regions. During solar proton events, the polar D region undergoes increased ionization, which in turn appears as phase advances on transpolar paths. The effect is well understood, and available models give corrections to the phase advances that return the Omega accuracies to their nominal 1 nautical mile (see, for example, Argo and Rothmuller, 1979). The auroral disturbances happen in the upper D region ( $>80$  km), and so VLF propagation is only marginally affected.



## 5.2 Predictions--Future

The need for HF communications and HF radars include real-time specification of the high-latitude conditions, e.g. the Q-index defining auroral oval size and location, and some method for positioning the main auroral/polar features such as arcs, the mid-latitude trough, the polar cusp, etc. Future capabilities should therefore include a Dynamics Explorer (DE) type satellite in a high inclination polar orbit that could in real time examine the polar and auroral regions in several optical wavelengths. From this information can be derived three-dimensional electron density profiles, and very accurate determination of polar/auroral feature location. Raytracing would allow (in near real time) the prediction of HF propagation parameters described in Table 2.

In addition, more rigorous specification of the polar/auroral morphology should be undertaken, with emphasis on locating auroral features under varying geomagnetic conditions. This would aid the communication planner (e.g., frequency allocation) in giving a statistical estimate of his frequency needs.

## 6. RECOMMENDATIONS

1. Attempt to better understand the basic physical processes governing:
  - a. Trough formation and disappearance
  - b. Auroral absorption (particle precipitation patterns)
  - c. Sporadic E dynamics
  - d. F-region anomalies (spread F, lacunae, etc.)
  - e. Cleft/cusp ionospheric dynamics
  - f. Ionosphere/neutral atmosphere interaction
2. Devise a better morphological description of the high-latitude ionosphere (appropriate ionospheric and geophysical parameters)
  - a. Main ionospheric trough (F region) polar orbiting satellites, Chatanika, EISCAT and Millstone incoherent scatter radars, observations of poleward edge
  - b. Auroral oval (sporadic and auroral E) polar orbiting satellites, DMSP, TIROS-N, DE, ionosonde network, incoherent scatter radars, IMS magnetometer chain, HF radars, TRANSIT satellite, GPS satellite, photometers
  - c. Cleft/cusp--UHF incoherent scatter radar (E- and F-region anomalies), ionosondes, advanced ionospheric sounders, polar orbiting satellites, spectrophotometric observations, HF reception on polar paths
  - d. Polar cap (E and F regions)--UHF incoherent scatter radar (E- and F-region anomalies), ionosondes, advanced ionospheric sounders, polar orbiting satellites
3. Development and verification of improved long-term and near-real-time models using
  - a. High quality research data now available (Chatanika radar, wideband satellite, Millstone radar)
  - b. The large synoptic data base existing (ionosonde network, all-sky camera network, magnetometers, riometers, HF backscatter, solar observations)

- c. Then develop better specification techniques (are there better indices for the separate features?)
4. Quantitative long-term and real-time predictions should be user-specific!
  - a. Operational frequency and fading rates for HF communications operators, ducting and non-great-circle propagation for HF sensing
  - b. Circulation and comparison of prediction techniques between international centers
  - c. Encourage better communication between scientists, predictors, and users.
5. Encourage funding agencies to support efforts to analyze and "synthesize" existing high-latitude data and to maintain existing data sources (e.g., ionosonde network).

## REFERENCES

[An asterisk (\*) indicates a paper in the 1979 Solar-Terrestrial Predictions Workshop Proceedings pertinent to high-latitude ionospheric predictions]

- \*Akinyan, S. T., I. M. Chertok, and E. M. Zhulina: Determination of PCA values by characteristics of solar radio bursts.
- \*Argo, P. E., and I. J. Rothmuller: Prophet: An application of propagation forecasting principles.
- \*Avdiushin, S. I., A. D. Danilov, A. B. Malishev, G. N. Novikova, and P. M. Svidskiy: Forecasting ionospheric and geomagnetic conditions in the forecasting center of IAG.
- Barghausen, A., J. Finney, L. Proctor, and L. Schultz (1969): Predicting long-term operational parameters of high frequency skywave telecommunications systems. Technical report ERL 110-ITS 78.
- \*Besprozvannaya, A. S., A. V. Shirochov, and T. I. Shchuka: On the approach to forecasting polar ionospheric conditions.
- \*Besprozvannaya, A. S., A. V. Shirochov, and T. I. Shchuka: On the diagnosis of the IMF parameters by the vertical ionosonde data from Vostok Station, Antarctica.
- Davies, R. M., Jr. (1972): The occurrence of spread F and its effect on HF propagation. Telecommunications Research Engineering Report 28.
- \*Goodman, J. M.: A resume of anticipated Fleetsatcom and Gapfiller scintillation effects during the peak of solar cycle 21 (1980-1982).
- \*Hatfield, V. E.: HF communications predictions 1978.
- \*Heckman, G. R.: A summary of the indices and predictions of the Space Environment Services Center.
- \*Hruska, J.: Forecasts of geomagnetic activity by Ottawa Magnetic Observatory: Their reliability and application.
- \*Hunsucker, R. D.: Morphology and phenomenology of the high-latitude E- and F-regions.

- \*Kerblay, T. S., E. M. Kovalevskaya, E. M. Zhulina, and L. M. Ishkova: Prediction of the characteristics of radio signals reflected from a horizontally-inhomogeneous ionosphere and the relevant requirements for prediction of ionospheric parameters.
- \*Klos, Z., and K. Stasiewicz: Estimation of solar flare influence on radio circuit transmission loss.
- \*Kovalevskaya, E. M., and E. M. Zhulina: On the statistical properties of the disturbed high-latitude ionosphere in radio wave propagation computations.
- \*Kuleshova, V. P., E. V. Lavrova, and L. N. Lyakhova: Forecasting of foF2 variations for ionospheric disturbances.
- \*Lassudrie-Duchesne, P., A. M. Bourdila, and H. Sizun: The French short-term radio propagation predictions in the decameter band.
- \*Maeda, R.: Radio propagation prediction services in Japan.
- \*Maeda, R., and H. Inuki: Radio disturbance warning issuance systems.
- \*Marubashi, K., Y. Miyamoto, T. Kidokoro, and T. Ishii: Forecasts of geomagnetic activities and HF radio propagation conditions made at Hiraïso, Japan.
- \*McNamara, L. F.: The use of ionospheric indices to make real- and near-real-time forecasts of foF2 around Australia.
- \*Meisel, D. D., B. Duke, and W. D. Savedoff: Minicomputer simulation of ionospheric radiowave propagation at decametric wavelengths.
- \*Mendillo, M., F. X. Lynch, and J. A. Klobuchar: Geomagnetic activity control of ionospheric variability.
- \*Mendillo, M., and J. A. Klobuchar: A morphology-based prediction scheme for the coupled latitudinal and local time development of F-region storms.
- Miller, D. C. and J. Gibbs (1975): Ionospheric analysis and modeling. Scientific Report 2, AFCRL-TR-75-0549.
- \*Mitra, S. N., and M. Sain: Short term prediction of ionospheric disturbances.
- \*Murthy, P. S. N., C. S. R. Rao, and M. Sain: Prediction of foF2 by the monthly ratio (MR) method.
- \*Ondoh, T. and T. Obu: Prediction of HF communication disturbances by pre-SC HF field increases on polar paths crossing the auroral zone.
- \*Oyinloye, J. O.: Prediction of radio wave absorption in the ionosphere.
- \*Shlionsky, A. G.: Prediction of waveguide propagation of radio waves using the extremal-parametric method on the basis of prediction of the key ionospheric parameters.
- \*Singleton, D. G.: Predicting transionospheric propagation conditions.



- \*Soicher, R.: Predictions of transionospheric signal time delays using correlative techniques.
- \*Tascione, T. F., T. W. Flattery, V. G. Patterson, J. A. Secan, and J. W. Taylor, Jr.: Ionospheric modeling at Air Force Global Weather Control.
- \*Thompson, R. L., and J. A. Secan: Geophysical forecasting at Air Force Global Weather Control.
- \*Troshichev, O. A., N. P. Dmitrieva, B. M. Kuznetsov, and V. P. Vasiliev: Interplanetary magnetic field and polar cap magnetic disturbances: Using the data for prediction of auroral electrojet activity.
- \*Turner, J. F., and P. J. Wilkinson: A weekly ionospheric index.
- \*Vanslette, K. E.: Communications predictions in the HF bands using local solar time.
- \*Vlasov, M. N.: Fundamentals of the physical forecast of the ionospheric plasma.
- \*Wilkinson, P. J.: Prediction limits for foF2.
- \*Zevakina, R. A., and M. V. Kiseleva: On the short-term prediction of the space-time distribution of the auroral absorption.
- \*Zevakina, R. A., and E. V. Lavrova: On the possibility of predicting variations in the F2-region parameters as a function of the IMF direction.

## APPENDIX A

### USES OF SATELLITE DATA

We envision the steps in understanding as proceeding from exploratory studies of phenomenology and morphology, through heuristic model construction, to definition of physical parameters. These parameters (electron, ion, and neutral densities; particle energy spectra; and energy sources and sinks) should support detailed calculations of either microscale or macroscale characteristics of the polar ionosphere, using rigorous physical equations.

Satellite measurements have contributed greatly to our initial steps in this process. Explorer XX, the Alouette and ISIS satellites have given us our first look at the general topside ionospheric features. The DMSPs have provided the coordinated picture of auroral luminosity and its relation to both top and bottomside profile perturbations. Injun and the Tiros satellites have examined energy deposition input. The geostationary satellites have monitored both magnetic intensity and particle parameters on field lines that base near the auroral oval. Although coverage is more appropriately described by the word "spotty" than by the word "comprehensive," the initial needs of the phenomenology and morphology requirements have been well satisfied.

To proceed further, there are two requirements: monitoring and converting the observational science to an experimental science. Since one of our main concerns is with electron density and its distribution (both radially and laterally), we must concern ourselves with means for monitoring this important information in the face of the extreme variability found in the polar regions. Remote sounding techniques are vital but must be extended by (1) providing angular resolution such as to separate distinct features (current technology senses laterally integrated densities over perhaps  $\pi$  steradians) and (2) by providing stationary platforms so that spatial and temporal variations may be decoupled. We urge strong support for the DE mission, and those parts of the "Open" program applicable to our mission.

This is the thrust of the monitoring requirement, but to proceed to a physical basis, we must take the steps that will convert an observational science to an experimental science. Three current examples of the ionospheric community's commitment to this approach come to mind: the heating experiments (Platteville and Arecibo), the chemical and water release experiments, and the electron-injection experiments (University of Minnesota type). We deplore the fragmentation of the shuttle "Amps" program, and urge the reinstatement of high-latitude ionospheric experimentation, such as can be pursued by extension of the shuttle "Wisp" program.

We foresee our "customers" as being broader than just the scientific community, with its interest in research and development and modeling; we must also serve the operational community. The operational community is engaged in assessing (in real time) the impact of geophysical variability on systems that are vital to the health and welfare of both man and his systems. By vigorously pursuing a program rooted in the aforementioned goals, we may best serve both communities.

## APPENDIX B

### PHYSICAL KNOWLEDGE OF THE PLASMA

#### B1. Aeronomical Models of Reactions

Taking into account that both the ionization by charged particles as well as UV solar radiation are important ionizing factors, and that the composition of the neutral atmosphere in the polar region is controlled by dynamic processes, the knowledge of the aeronomy of the weakly ionized plasma is important. The open questions are: (1) aeronomical processes under different states of the neutral atmosphere; (2) reaction constants dependence on thermospheric temperature (taking into consideration the broad range of temperature); and (3) the role of X-rays generated by bremsstrahlung processes in the determination of chemical composition.

#### B2. Precipitating Particle Interaction with the Ambient Plasmas

The principal mechanism in the interaction of high energy particles with the neutral atmosphere and ambient plasma is known. Further investigation must be undertaken into (1) binary particle interactions (elastic and

inelastic) and (2) collective particle interactions (including instabilities).

Still more controlled determination of the collision cross section in the broad range of energies is necessary to specify the collisions with neutral molecules. The mechanisms for plasma instability generation by precipitating particles and their influence on particle spectra have to be investigated. Finally, the distribution function of the particles at several levels in the E and F regions must be calculated as the first step for ionization rate and particle transport calculations.

### B3. Small-scale Structures in Plasma and Their Generation Mechanisms

Small-scale irregularities in the polar E- and F-region ionospheric plasma have to be further investigated. They have an important role in HF as well as VHF propagation. Most aspects of the generation mechanism of irregularities in the E region, particularly in the polar electrojet, are known, but there still are open questions about the F region, especially the mechanism of spread F creation. The range of factors that can create instability--particle precipitation, wind shear,  $E \times B$  and gradient drift instability--have to be considered. The nonlinear calculations of these mechanisms, which permits the determination of the fluctuation spectrum, have yet to be done.

### B4. Average Plasma Transport Parameters

It is equally important to investigate the influence of small-scale irregularities in the ionospheric plasma on average plasma transport parameters, particularly in the polar E region. Such transport plays an important role in many problems of ionospheric dynamics such as the configuration of the three-dimensional ionospheric-magnetospheric current system, partition of the field-aligned current into the ionospheric east-west Hall current and north-south Pedersen current, and Joule heating caused by the electrojet. As the irregular structure is caused by various instability types, the additional transport caused by irregularities in different situations has to be analyzed, and its role in the determination of the plasma state must be evaluated. Calculations and the experimental verification must be done not only in connection with field-aligned currents, but also with perpendicular currents, as most of the instability mechanisms create motion across the field lines.



# ON THE APPROACH TO FORECASTING POLAR IONOSPHERIC CONDITIONS

A. S. Besprozvannaya, A. V. Shirochkov, and T. I. Shchuka  
The Arctic and Antarctic Research Institute  
Leningrad, USSR

Results of an extensive analysis of the major properties of polar ionospheric main anomalous events are summarized. The monitoring of large-scale features of the ionization distribution that are the projections of large-scale structural characteristics of magnetospheric plasma on the upper ionosphere is suggested as a basic principle of polar ionospheric condition forecasting. It is concluded that the processes of the magnetosphere/ionosphere interaction appear to play a predominant role in the creation of the polar ionosphere.

## 1. INTRODUCTION

Recently the physics of the polar ionosphere has become one of the major studies in geophysical research. Significant results have been obtained, increasing our knowledge of general processes responsible for the ionization balance in the polar ionosphere. This has become possible primarily due to extensive use of satellite and rocket data.

Satellite data provide better understanding of earlier ground-based observational data, and allow the researchers to reassess the information they got from the network of ground-based ionospheric stations and to use it for the solution of today's problems of the polar ionosphere. As a result of such studies, a very important conclusion has been made, that the processes of the magnetosphere/ionosphere interaction appear to play a predominant role in the creation of the polar ionosphere.

Therefore, we believe that those who are concerned with the high-latitude ionosphere should give up the traditional forecasting methods used for mid-latitude ionosphere forecasting, and instead use as the main principle the monitoring of large-scale peculiarities of the ionization distribution, which are the traces of mapping of large-scale structural parameters of magnetospheric plasma on the upper ionosphere.

## 2. DATA ANALYSIS

### 2.1 F-region

The spatial distribution of the electron concentration in the F2-region maximum for a high-latitude area is shown in Figure 1. The hourly median values of the F2-region critical frequencies (foF2) from the global ionosonde network are used on the plots of foF2 isolines for 00 hours UT in December 1958 (solar maximum) and 1964 (solar minimum). One can easily see that the main features of foF2 spatial distribution in Figure 1 closely correspond to similar data from satellite measurements (Nishida, 1967; Marubashi, 1970; and Whitteker et al., 1978). They are:

a. A region of low values of foF2 on the night side in mid-latitudes, the so-called ionospheric trough.

b. A region of high values of foF2 on the night side poleward from the ionospheric trough. This is the auroral peak, particularly apparent in the years of maximum solar activity.

c. A region of high values of foF2 in the vicinity of the geomagnetic pole. The polar peak is seen as a tongue-shaped deformation of the daytime foF2 isolines near the geomagnetic pole.

Let us consider these features in greater detail. An ionospheric trough is a kind of demarcation line between the high- and mid-latitude ionosphere. The equatorward wall of it is under the influence of the processes typical of the mid-latitude ionosphere. On the other hand, the poleward wall of the ionospheric trough is formed by the processes connected with the precipitation of the auroral electrons. The shape of the trough changes drastically with the solar cycle (Besprozvannaya, 1975b). In years of higher solar activity it appears as a region of wide depression, while during lower solar activity it forms a narrow gap with a rapid increase of foF2 values on either side (Figure 2). In the trough minimum in winter foF2 values drop as low as 20-25 percent from the normal daytime values at all levels of solar activity. The maximum electron density in the center of the trough appears to be approximately  $1.5 \times 10^5 \text{ cm}^{-3}$  in the years of high solar activity and equal to or less than  $1.2 \times 10^4 \text{ cm}^{-3}$  in the years of low solar activity. The position of the center of the trough is determined mainly by the motion of the poleward wall of the trough, depending on the displacements of the low-energy electron precipitation boundary (Turunen and Liszka, 1972; Halipov et al., 1977). The trough position has been shown to depend on geomagnetic disturbance level (evaluated in Kp) as well as on the ring current intensity (evaluated in Dst) (Rycroft and Byrnell, 1970; Zaitseva and Alexeyva, 1975).

The ionization increasing from the trough toward the pole can be treated as the manifestation on the F2-region heights of both the day and night parts of the plasma ring, which is a region of enhanced ionization revealed by the satellite data from the ionosphere top soundings (Thomas and Andrews, 1969). The plasma ring is stationary with respect to the sun and its position in space changes with variations in geomagnetic activity. Its equatorial boundary at the night side is located at  $65^\circ$ - $68^\circ$  geomagnetic latitude, while at the day side it is  $75^\circ$ - $80^\circ$  geomagnetic latitude when the geomagnetic field is quiet. An increase of geomagnetic activity brings about significant changes in the positions of both polar and auroral peaks (Marubashi, 1970; Sato and Colin, 1960). As a result large variability in the F2-region electron concentration is observed (Figure 3).



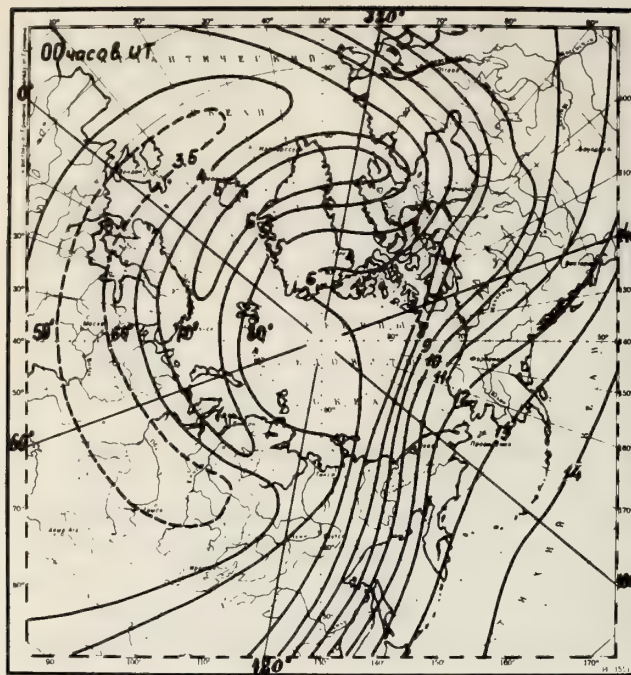


Figure 1a. foF2 median values (MHz) for 00 UT in December 1958.

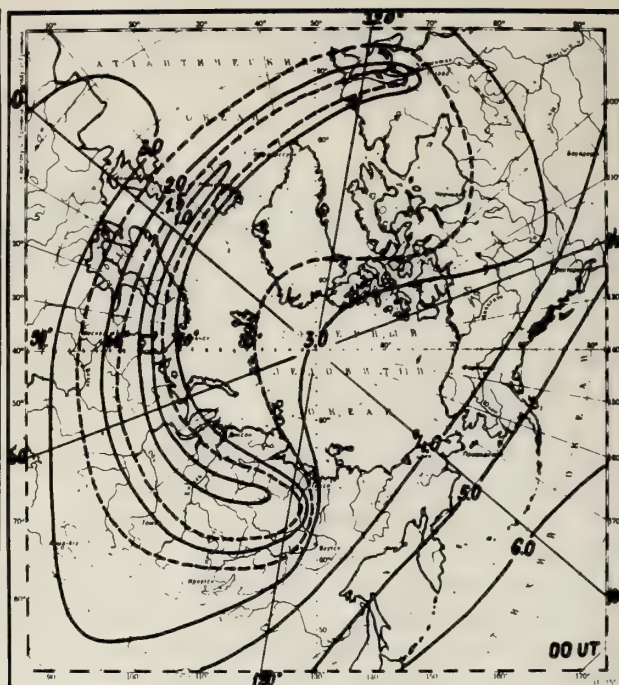


Figure 1b. foF2 median values (MHz) for 00 UT in December 1964.

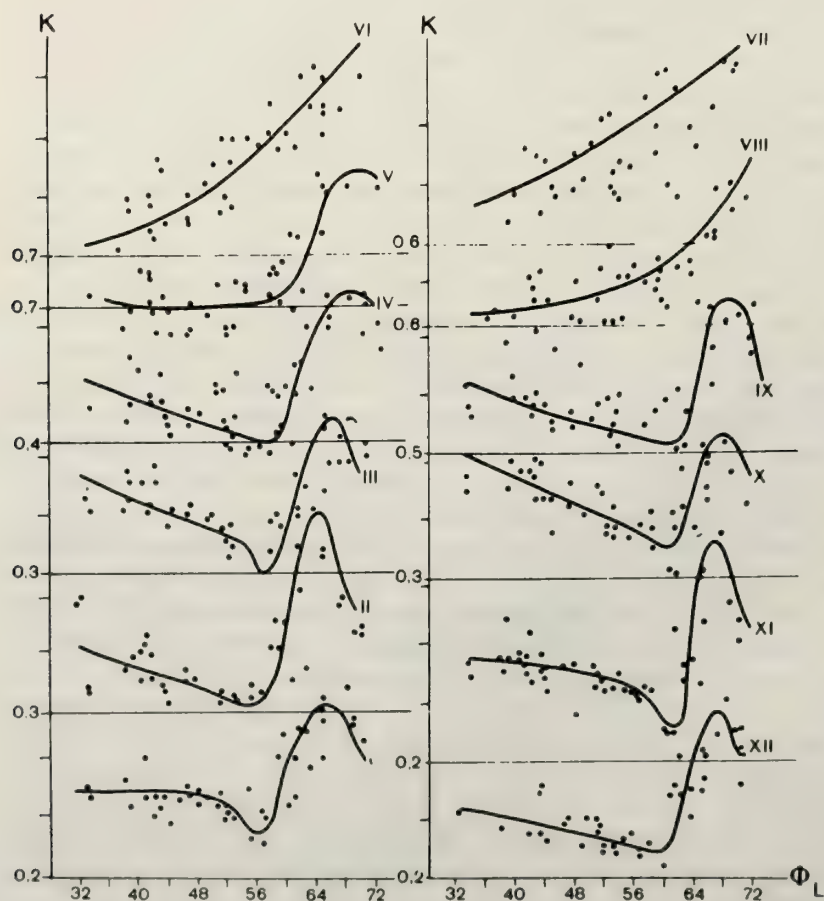


Figure 2a. Latitude dependence of the normalized-to-noon values of critical frequency of the F2 region at 02 LT (K) for 1958.



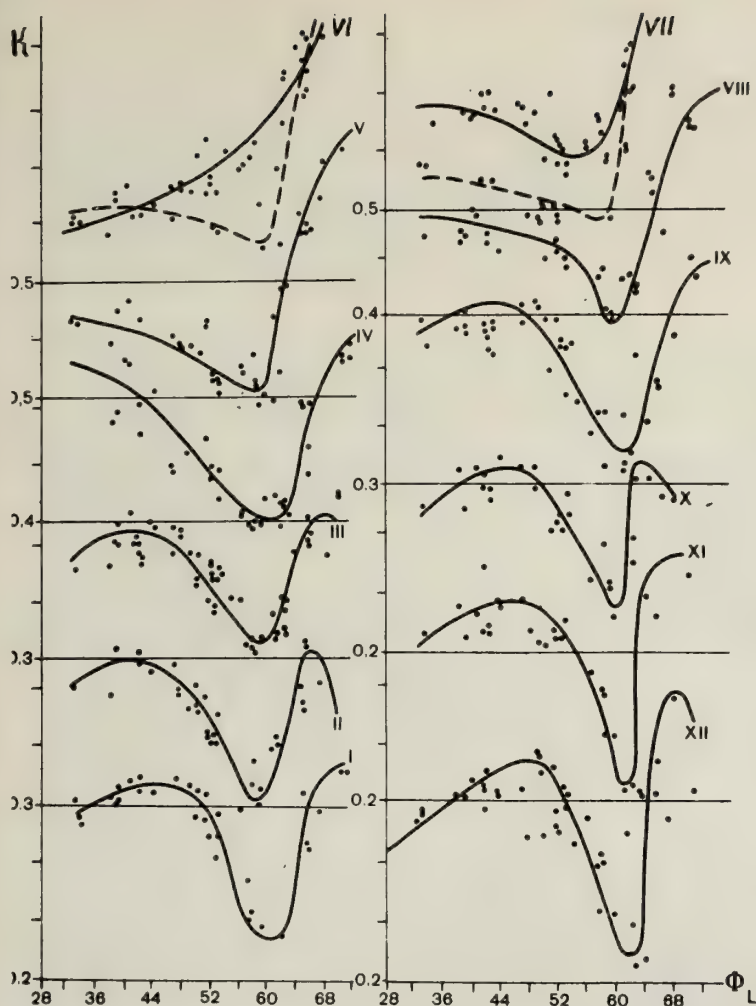


Figure 2b. Latitude dependence of the normalized-to-noon values of critical frequency of the F2 region at 02 LT (K) for 1964.

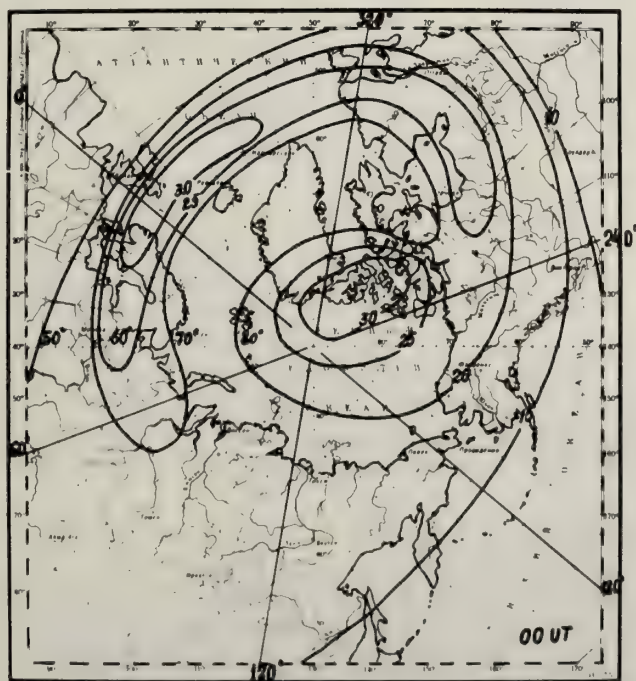


Figure 3. foF2 variability coefficient ( $v_n = \sigma/foF2_{med}$ ) isolines (in percent) for 00 UT in December 1958.

## 2.2 E Region

The same global ionosonde data were used to plot the spatial/temporal distribution of sporadic E ionization in high latitudes for the winter months of high and low solar activity. The results of the analysis are given for two major classes of Es forms: the so-called auroral E, which includes the traces from the night E region and Es layers with the group retardation at the high-frequency end of the trace; and other types of Es layers, including "f," "l," "c," "a," and "h" types, in accordance with the URSI recommendations (Besprozvannaya et al., 1978).

Figure 4 shows the corresponding plots depicting Es ionization distribution for quiet and disturbed geomagnetic conditions under the high and low solar activity conditions. Analyzing these data, the following conclusions can be made:

a. Under quiet geomagnetic conditions a region of enhanced Es occurrence is located in relatively high geomagnetic latitudes, including the polar cap area; solar cycle variations are negligible except for a rather small shift of the region towards the equator about midnight in 1963-64.

b. The pattern changes significantly with an increase in geomagnetic activity: (1) the ring zone of enhancement of Es occurrence is asymmetric relative to the geomagnetic pole. The whole of the region is displaced to lower latitudes. The equatorial boundary of this region at the night side is at  $60^{\circ}$ - $62^{\circ}$  of geomagnetic latitude while the poleward boundary is at approximately  $72^{\circ}$ - $73^{\circ}$ . With decreasing solar activity, the region moves to higher latitudes; the equatorial boundary is at  $64^{\circ}$ - $67^{\circ}$  latitude and the poleward one at  $74^{\circ}$ - $78^{\circ}$ . (2) The region of maximum Es occurrence is divided into two subzones. This phenomenon is especially apparent in the years of high solar activity when a narrow band with very low Es occurrence at  $65^{\circ}$ - $68^{\circ}$  latitude is formed. On both sides of this band the Es occurrence sharply increases, reaching its maximum values at latitudes  $63^{\circ}$ - $64^{\circ}$  and  $68^{\circ}$ - $70^{\circ}$ , respectively. In the years of low solar activity such division of the Es occurrence region embraces evening hours, the depth of the trough being smaller.

The positions of the auroral oval for quiet and disturbed conditions (Feldstein and Starkov, 1967) together with the positions of the boundary of the 1.3 keV electron precipitation (Starkov, 1974) are shown in Figure 4. It is evident that under disturbed conditions the Es enhanced occurrence region coincides with both zones.

The distribution of auroral E is shown in Figure 5, in plots similar to Figure 4, for different levels of solar activity and geomagnetic disturbance. The important features of these data are:

a. There is no significant dependence on the level of the geomagnetic activity. In spite of its change, the region of the auroral E occurrence has been found to have a horseshoe-like form, open to the day side with a similar probability of occurrence. Nevertheless, the minor displacements of the auroral E occurrence region due to geomagnetic activity variations are observed. The general trend for both equatorial and poleward boundaries of this region is towards the lower latitudes caused by the geomagnetic activity increase. These displacements do not exceed  $2^{\circ}$  to  $3^{\circ}$ .

b. Solar activity variations cause apparent changes in the size and location of the region considered, mainly with quiet geomagnetic conditions. During low solar activity, the region diminishes greatly in width and extent.

c. The division of the auroral E occurrence region is observed for periods of high geomagnetic activity. The phenomenon is more significant in the



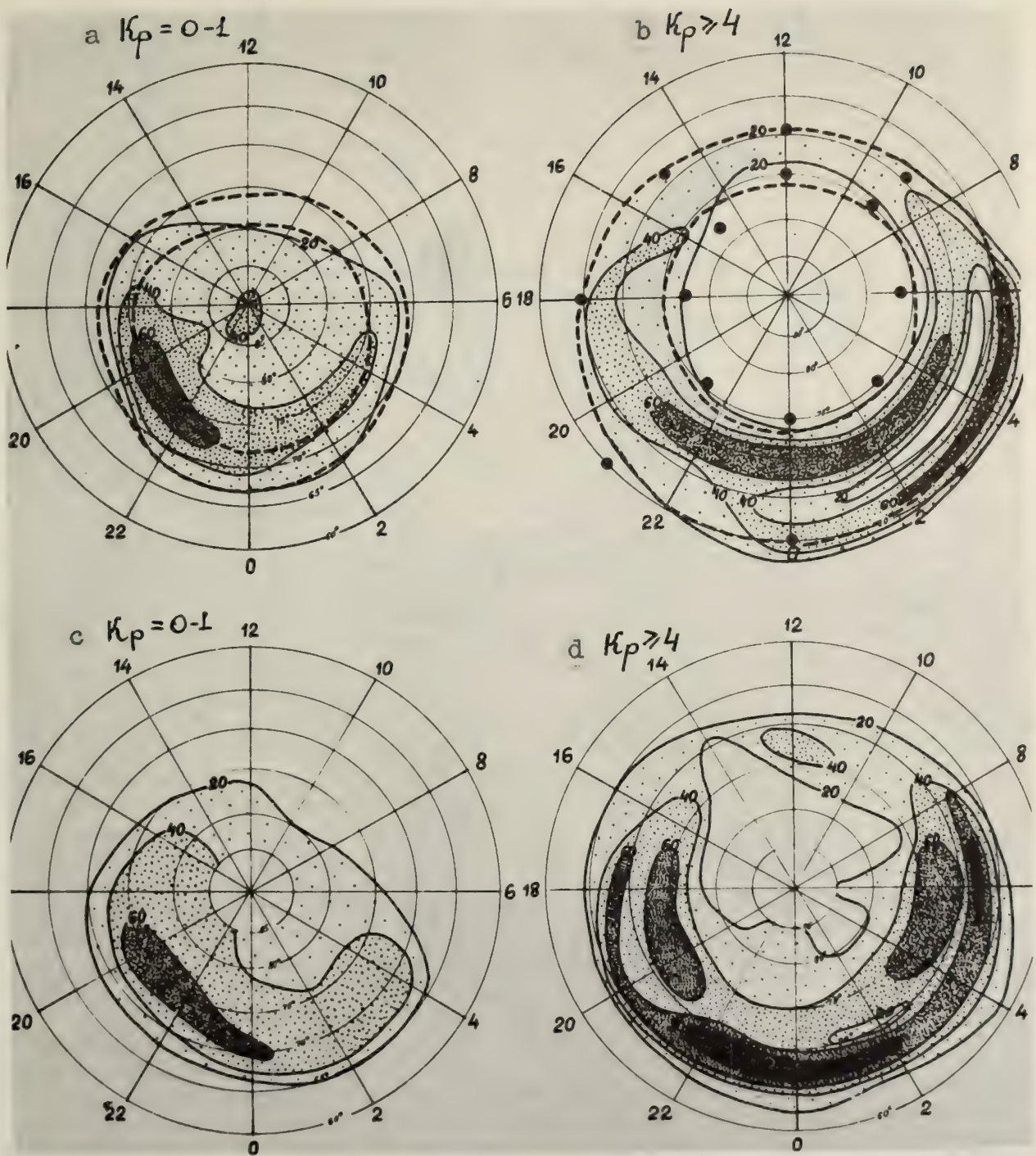


Figure 4. Polar plots of the probability of sporadic E, winter 1958 and 1963-64. Corrected geomagnetic latitude and time. Dashed line is auroral oval and filled circles are low energy electron precipitation boundaries (Starkov and Feldstein, 1967; Starkov, 1974).



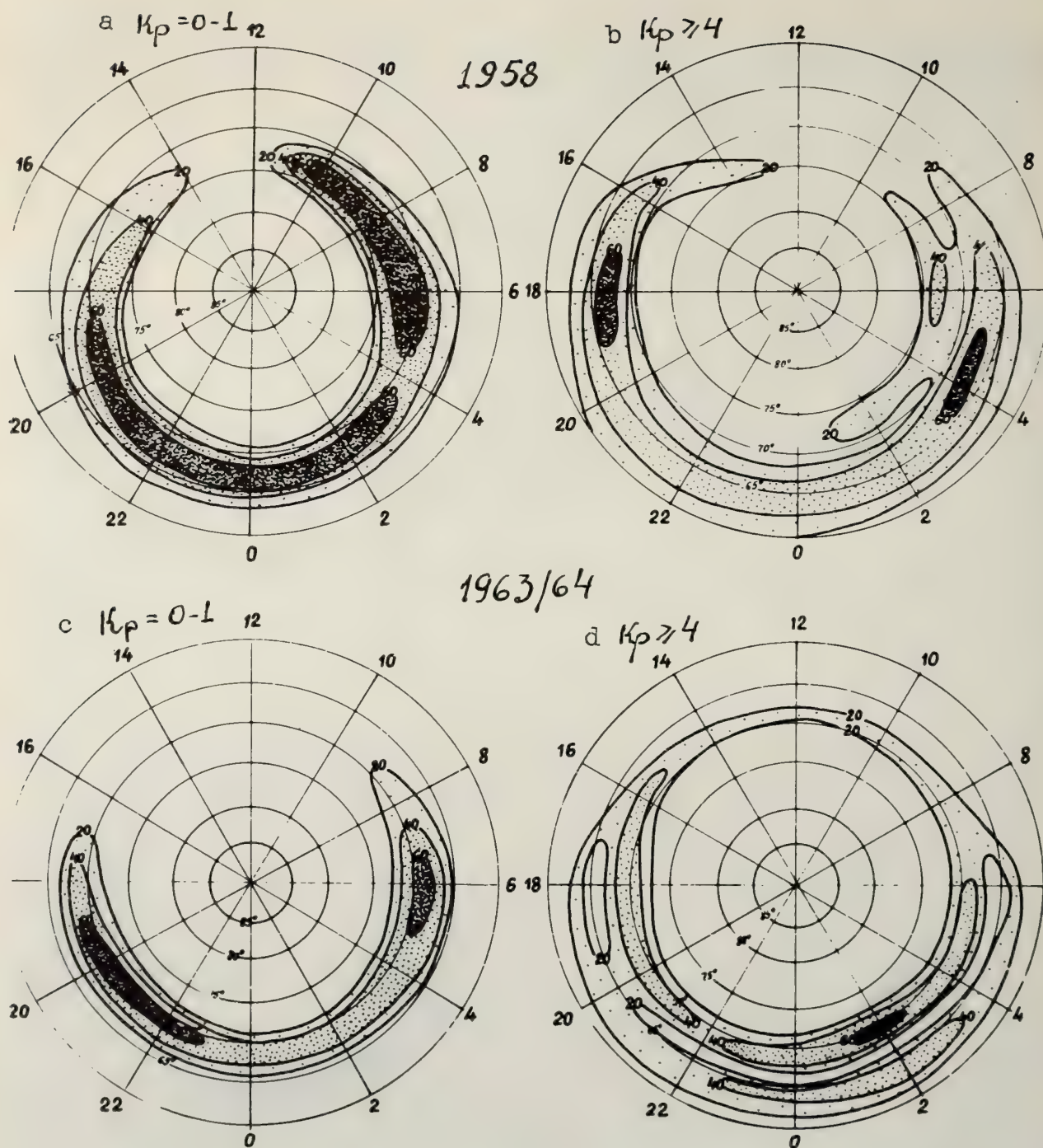


Figure 5. Polar plots of occurrence probability of auroral E, winter 1958 and 1963-64. Corrected geomagnetic latitude and time.

epochs of low solar activity (1963-64). The trough with auroral E occurrence lower than 20 percent is located at  $\phi' \approx 67^\circ$ , while the maxima of auroral E occurrence is observed at  $\phi' \approx 63^\circ$  and  $\phi' \approx 70^\circ$ , respectively.

There is no consensus yet concerning the nature of the auroral E. Nevertheless, several studies of the correlation between the proton aurorae and sporadic E ionization (Eather and Jacka, 1966; Evlashin and Nekhludova, 1971; Gorely and Pirog, 1978) give a valid basis for suggesting that the nature of the auroral E is determined by the fluxes of precipitating low-energy protons.

### 2.3 D Region

Riometer network data obtained over the Soviet Arctic for a period of many years allow us to construct polar diagrams of the auroral absorption (AA) occurrence for winter months at different levels of solar activity as well as geomagnetic disturbance (Shchuka, 1978). One of the plots of this type for the winter of 1964 (solar minimum) is shown in Figure 6. The probability of AA occurrence was taken as the ratio of the number of events with the absorption value higher or equal to 0.5 db (at 30 MHz) to the complete period of observation.

Under quiet geomagnetic conditions a narrow ring zone of 10 percent AA occurrence is located along  $\phi \approx 70^\circ$ . With an increase of geomagnetic activity ( $K_p = 2$ ) two AA enhancements (30 percent) occur at 02-04 UT and 10 UT with the AA occurrence region expansion towards lower latitudes ( $\phi \approx 62^\circ$ - $63^\circ$  in the morning hours). When  $K_p = 3$ , AA occurrence increases significantly with an after-midnight maximum, rising to 70 percent, closer to 00 hours GLT and expanding with time. Simultaneously the noon maximum becomes 50 percent. The 10 percent isoline is located at  $61^\circ$ - $62^\circ$  almost all day. Some poleward expansion of the whole AA region is noticeable in the morning as well as in the noon hours.

With  $K_p$  as high as 4, three maxima of AA enhancement (70 percent) are observed at 00-02, 06 and 10 hours GLT. AA occurrence equal to 60 percent is recorded from 22 to 13 hours GLT continuously. On the other hand, with all levels of geomagnetic disturbance there is a permanent deep minimum in AA occurrence (less than 20 percent) at about 18 hours GLT.

During the epoch of high solar activity the general features of AA distribution are similar to those described above. It should be added only that there is some movement of the equatorial boundary of the AA region towards lower latitudes and a less significant displacement of its poleward boundary. The positions of maxima in the AA occurrence area shift towards the equator (for  $1^\circ$ - $2^\circ$  of latitude).

The same Arctic and Antarctic riometer data allow us to study the main large-scale features of such spectacular high-latitude geophysical phenomena as the polar cap absorption (PCA) events (Driatsky, 1974). The study of the latitudinal distribution of the absorption is worthwhile; it is from these data that we can learn a great deal about the real cut-off rigidity of the solar cosmic ray (SCR). The day-to-night ratio of absorption is also a very important parameter permitting us to get valuable information concerning aeronomical processes in the polar lower ionosphere. The shape of the latitudinal distribution of absorption in the North and South Hemispheres reflect both the influence of the interplanetary magnetic field (IMF) on SCR distribution and the character of SCR entry into the magnetosphere. The presence or absence of the midday recovery of the absorption effect can give some information on



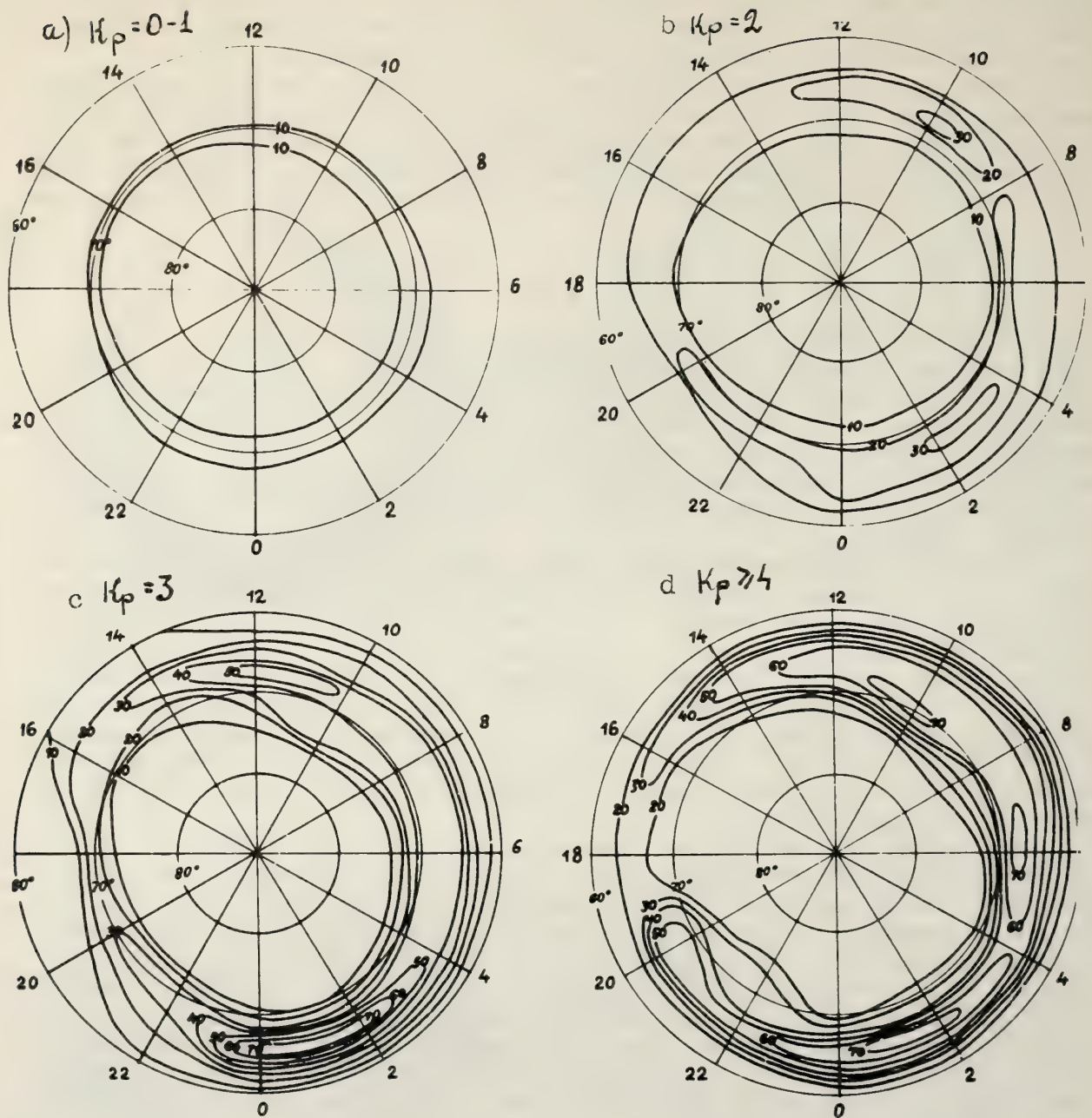


Figure 6. Polar plots of the probability of auroral absorption (from 30 MHz riometer data) occurrence for values  $\geq 0.5$  db in winter of 1964 in terms of corrected geomagnetic latitude and time.

important properties of the SCR. The riometer data allow us to get detailed information about the influence of SC magnetic storms on the spatial/temporal distribution of the absorption during solar proton events.

On the basis of the ground-based ionospheric observation data detailed catalogues of PCA events are prepared for the period 1938-1978 (Besprozvannaya, 1962; Driatsky, 1974; and others). This work is continued. Some tech-



niques for the computation of electron density profiles during PCA events on the basis of multi-frequency riometer data are proposed together with some original methods of calculating ionization rates for SCR accounting for their energy loss during their penetration into the atmosphere.

### 3. DISCUSSION

One can state definitely now that the large-scale features of anomalous phenomena in the polar ionosphere are closely related to the structural parameters of the magnetospheric plasma.

The ionosphere/magnetosphere interaction is realized via three main channels: interchange of the thermalized plasma between the ionosphere and magnetosphere; precipitation of energetic particles of the magnetospheric plasma into the ionosphere; and a mutual penetration of electric fields, generated both in the ionosphere and magnetosphere. Hence there is an excellent opportunity to use the ground-based ionospheric observations to solve direct and reverse problems: to determine the main features of the polar ionosphere by means of the information on magnetospheric plasma conditions, and to use numerous ionospheric data for the diagnosis of the magnetospheric parameters.

The most interesting features of such interaction are discussed below.

a. The statistical relationship between the position of the main ionospheric trough in the F2 region and the positions of the plasmopause (Besprozvannaya, 1975a; Grebovsky et al., 1976; Rycroft and Burnall, 1970) and DR current. As an illustration of this relationship, Figure 7 shows an average position of the ionospheric trough for 25 global magnetic storms in 1958 (solid line). Separate dots depict a trough position for a storm. The position of DR current derived from the satellite data is shown by the dashed curve (Zaitseva and Alexeyeva, 1975). These data confirm the possibility of predicting the DR current position from the ionosonde data.

b. The relationship between the boundary of the enhanced ionization in the ionosphere and the soft electron precipitation. Up to now the two main sources of precipitating particles were discovered by means of satellite experiments such as (1) thermalized plasma of the magnetosheath penetrating into the high-latitude atmosphere on the day side in the area of the polar cusp, and (2) auroral plasma, originating in the plasmasheath of the magnetosphere tail. The former source is assumed as responsible for the day part of the plasma ring (the polar peak) and the latter for the night part of the plasma ring (the auroral peak). Figure 8 shows the location of the equatorial boundary of F2-region anomalous ionization (Besprozvannaya, 1972) and the one for the plasma ring from the satellite measurements (Andrews and Thomas, 1969). A close coincidence of both curves demonstrates the evident opportunity of using the ionosonde data of the anomalous ionization latitudinal distribution for determining the equatorial boundary of the soft electron precipitation region.

c. The statistical relationship between the location of the auroral E occurrence equatorial boundary and the plasmopause position. On the basis of the information available at present it is believed that there might exist a relationship between the auroral E ionization and DR-current proton precipitation (Aarness et al., 1970; Bernstein et al., 1974; Fukunishi, 1975;

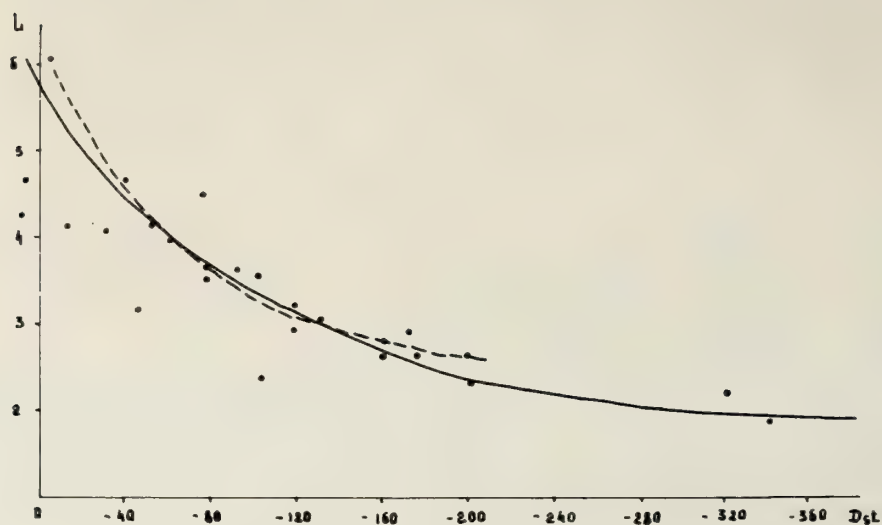


Figure 7. Average position of the ionospheric trough minimum in the night sector for 25 global magnetic storms versus the  $D_{st}$ -index (solid line). The dots depict individual values. The dashed line shows the statistical position of the DR current by satellite measurements (from Zaitseva and Alexeyeva, 1975).

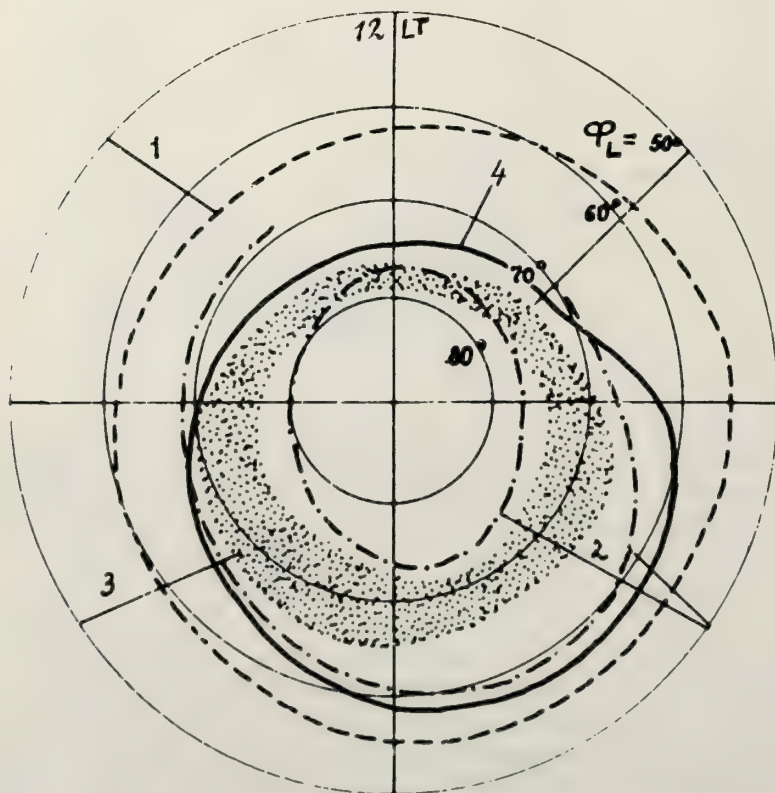


Figure 8. The positions of the equatorial boundaries of the F2-region anomalous ionization (4) (Besprozvannaya, 1974). The Thomas's plasma ring (2), the position of the plasmopause (1), and the auroral oval (3) are also shown (Andrews and Thomas, 1969).



Mizera, 1974). Figure 9 shows positions of the equatorial boundaries of the auroral E occurrence region and the plasmopause location (Bernstein et al., 1974). These data prove the possibility of using sporadic E ionization data for the evaluation of statistical position of both the plasmopause and the DR current.

d. The control of the sporadic E ionization inside the polar cap by the IMF. Numerous satellite measurements demonstrate clearly that the total amount of energy brought by the precipitating particles into the polar cap ionosphere would equal approximately 1 percent of the energy brought into the auroral ionosphere (Heikkila, 1972; Hultqvist, 1972). Therefore the Es formation in the polar cap is affected by a number of dynamic factors, which are found to be under the strong influence of the IMF (Heppner, 1972; Lyatzkaya et al., 1978). It was found that favorable IMF orientation for Es occurrence in the Northern Hemisphere is  $B_y > 0$ , while in the Southern Hemisphere it is  $B_y < 0$ . The relationship can be used for the IMF polarity diagnosis. (A paper by the present authors dealing with this problem in detail is now in press.)

e. The dependence of the latitudinal distribution of the absorption during the PCA events on the IMF polarity. There are some valid reasons to believe that the sign change of the azimuthal component of the IMF causes the corresponding change in the sign of the north-south asymmetry of the PCA latitudinal distribution (Shirochkov, 1977). In the years of stable IMF polarities one can expect larger magnitudes of the PCA absorption in the Northern Hemisphere with  $B_y > 0$ , while in the Southern Hemisphere,  $B_y < 0$ . It has been shown that the latitudinal chains of riometer data are more useful for the analysis of all the features corresponding to the IMF polarity change during the PCA events than fast-moving satellite data. In that case the ground-based observational data allow us to evaluate the time of the IMF sectoral structure boundary passing across the Earth.

f. The connection of the midday recovery of the absorption during the PCA events caused by the proton flares with a "soft" proton spectrum. Therefore, the information on the midday recovery effect can be used for evaluating proton spectra shape as well as for speculating on the solar flare coordinates (Ulyev, 1978).

#### 4. CONCLUSIONS

The results of this research allow us to conclude that the behavior of the polar ionosphere main anomalies can be considered as statistical models of the magnetospheric plasma structural peculiarities. Therefore, the scientific rationale for the development of ionospheric networks must be based on the concept that ionospheric stations should be regarded not only as simple data collectors but as efficient and useful devices for geophysical research of the behavior of magnetospheric and solar plasma.

A principal scheme of high-latitude ionospheric forecasting is to provide continuous control of the polar ionosphere "hot points" where the direct intrusion of the magnetospheric plasma occurs. Such hot points are located in the belt of  $75^{\circ}$ - $80^{\circ}$  of invariant latitude on the noon geomagnetic meridian (a plasma intrusion through the polar cusp) and the region of  $60^{\circ}$ - $65^{\circ}$  of invariant latitude (an auroral plasma intrusion from the plasmasheath).



Such a network of ionospheric stations allows us to use them most efficiently for the diagnosis and prediction of ionospheric conditions as well as magnetospheric parameters.

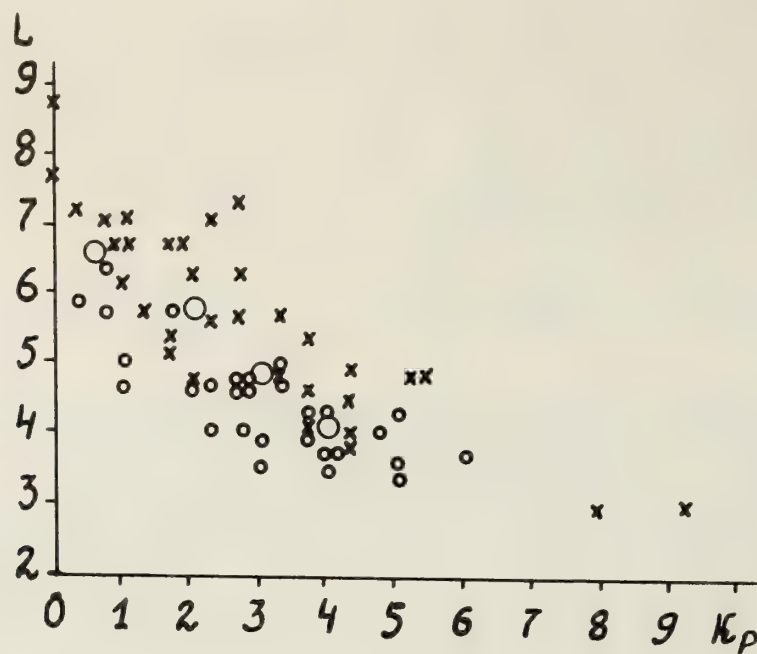


Figure 9. The positions of the plasmapause (crosses) and the equatorial boundary of the DR current (small circles) as a function of geomagnetic activity (Bernstein et al., 1974). Large circles depict the equatorial boundary of the auroral E occurrence region.

#### REFERENCES

- Aarnes, K., R. Amundsen, H. R. Lindalen and F. Soraas (1970): Proton precipitation in noon-midnight meridian and its relation to geomagnetic activity. Phys. Norvegica, 4:73.
- Andrews, M. K., and J. O. Thomas (1969): Electron density distribution above winter pole. Nature, 221:223.
- Besprozvannaya, A. S. (1962): Abnormal polar cap absorption associated with strong chromospheric flares on the Sun for the period of 1938-1959. Journ. Phys. Soc. Japan, 17:146.
- Besprozvannaya, A. S. (1972): Spatial-temporal distribution of F2 region anomalous ionization in high latitudes. Proc. Arctic & Antarctic Research Inst., 310:66. (In Russian).
- Besprozvannaya, A. S. (1974): The morphology of the high latitude F2 ionization distribution. In High-latitude geophysical events. Leningrad, Nauka, pp. 150-178. (In Russian).

- Besprozvannaya, A. S. (1975a): Connection between the ionospheric trough and the plasmopause in the afternoon sector. The Problems of Arctic and Antarctic, 45:73. (In Russian).
- Besprozvannaya, A. S. (1975b): The global distribution of night ionization in the F2 region maximum by the ground-based ionosonde data. Proc. of the AARI, 322:185. (In Russian).
- Besprozvannaya, A. S., A. V. Shirochkov, and T. I. Shchuka (1978): The dynamics of the high latitude ionospheric E region. To appear in The Journ. of Atmosph. Terr. Phys.
- Bernstein, W., B. Hultqvist, and H. Borg (1974): Some implications of low latitude observations of isotropic precipitation of ring current protons beyond the plasmopause. Planet. Space Sci., 22:767.
- Driatsky, V. M. (1974): Nature of anomalous absorption of cosmic radio emissions in the lower ionosphere of high latitudes. Leningrad, Gidrometeoizdat, 224 p. (In Russian).
- Eather, R. H., and F. Jacka (1966): Auroral hydrogen emission. Austr. Journ. Phys., 19:241.
- Evlashin, L. S., and Z. I. Nekhludova (1971): Polar auroral and sporadic ionization in the E region at Murmansk. In: Geophysical phenomena in the auroral zone, Leningrad, Nauka, p. 69-74. (In Russian).
- Feldstein, Y. I., and G. V. Starkov (1967): Dynamics of auroral belt and polar geomagnetic disturbances. Planet. Space Sci., 15:209.
- Fukunishi, H. (1975): Dynamic relationship between proton and electron substorms. Jour. Geophys. Res., 80:553.
- Gorely, K. I., and O. M. Pirog (1978): Observation of hydrogen emission and sporadic ionization of the E region. Geomagn. and Aeronomy, 18:354.
- Grebowsky, J. H., N. G. Maynard, Y. K. Tulunay, and L. Lanzerotti (1976): Coincident observations of ionospheric trough and the equatorial plasmopause. Planet. Space Sci., 26:1177.
- Halipov, V. L., Yu. I. Galperin, Yu. V. Lisakov, K. Cranier, L. M. Nikolayenko, and J. Souvour (1977): The diffuse auroral zone: The formation and dynamics of the poleward boundary of the ionospheric trough in the evening sector. Space Investigations, 15:708. (In Russian).
- Heikkila, W. J. (1972): Penetration of particles into the polar cap and auroral regions. In: Critical problems of magnetospheric physics. Proc. Joint COSPAR/IAGA/URSI Symp., Madrid, p. 67.
- Heppner, J. P. (1972): Polar cap electric field distributions related to the interplanetary magnetic field direction. Journ. Geophys. Res., 77:4877.
- Hultqvist, B. (1972): Auroral particles. In: Cosmical Geophysics, Oslo, 161-79.

- Lyatskaya, A. M., V. B. Lyatsky, and Yu. P. Maltsev (1978): Influence of longitude currents on the profile of electron concentration. Geomagn. and Aeronomy, 18:229. (In Russian).
- Marubashi, L. (1970): Structure of the topside ionosphere in high latitudes. Journ. Radio Res. Labs., 17:335.
- Mizera, P. F. (1974): Observations of precipitating protons with ring current energies. Journ. Geophys. Res., 79:581.
- Nishida, A. (1967): An average structure and storm-time change of the polar topside ionosphere and sunspot minimum. Journ. Geophys. Res., 72:6051.
- Rycroft, M. J., and S. J. Burnell (1970): Statistical analysis of movements of the ionospheric trough and the plasmopause. Journ. Geophys. Res., 75:5600.
- Sato, T., and L. Colin (1969): Morphology of electron concentration enhancement at a height of 1000 km at polar latitudes. Journ. Geophys. Res., 74:2193.
- Shchuka, T. I. (1978): The dependence of the spatial-temporal distribution of auroral adsorption on magnetic activity level. In press. (In Russian).
- Shirochkov, A. V. (1977): North-South assymetry in the intensity of absorption of PCA type. Geomagn. and Aeronomy, 17:445. (In Russian).
- Starkov, G. V. (1974): Planetary morphology of the aurorae. Polar Aurorae, 19:5.
- Thomas, J. O. and M. K. Andrews (1969): The transpolar exospheric plasma. 3. A united picture. Planet. Space Sci., 17:433.
- Turunen, F., and L. Liska (1972): Comparison of simultaneous satellite measurements of auroral particle precipitation with bottomside ionosonde measurements of the electron density in the F region. Journ. Atmosph. Terr. Phys., 34:365.
- Ulyev, V. A. (1978): Relation of midday recovery effect onset during the PCA to flare solar longitude, Proc. Arctic and Antarctic Research Inst., 350:73. (In Russian).
- Whitaker, J. H., G. G. Shepherd, C. D. Anger, J. R. Burrows, D. D. Wallis, D. M. Klumpar, and J. K. Walker (1978): The winter polar ionosphere. Journ. Geophys. Res., 83:1503.
- Zaitseva, S. A., and N. Ya. Alexeyeva (1975): Energy dissipation of DR-belt and the condition in ionosphere. In: Substorms and Magnetospheric Disturbances, Leningrad, Nauka, 156-170. (In Russian).



# MORPHOLOGY AND PHENOMENOLOGY OF THE HIGH-LATITUDE E- AND F-REGIONS

Robert D. Hunsucker  
Geophysical Institute  
University of Alaska  
Fairbanks, Alaska 99701

This paper presents results obtained at high-latitude observatories in the last twenty years on the behavior of E- and F-region ionization. A bibliography on the subject is also presented.

Behavior of E- and F-region ionization during day and night for quiet and disturbed conditions in the auroral and polar regions is described. Daily, seasonal and sunspot variations are also outlined.

## 1. INTRODUCTION

The purpose of this paper is to present representative data obtained by various radio techniques in the last twenty-five years on the morphology of ionization in the high-latitude E- and F-regions. A brief bibliography of pertinent publications on this topic is also included, concentrating on literature published in the last fifteen years. It is the intent of the author to provide background information which will be of use to geophysicists and radio engineers in the development of high-latitude ionospheric predictions.

Review papers and/or bibliographies on high-latitude ionosphere morphology and propagation are relatively rare and those available at the present time are by Agg [1957], Penndorf and Coroniti [1959], Kirby and Little [1960], Gassman [1963], Landmark [1964], URSI [1964], Hunsucker [1967], Folkestad [1968], Hunsucker and Bates [1969], and Bates and Hunsucker [1974].

## II. DATA BASE AND DISCUSSION OF SPECIFIC EVENTS

### A. E-Region

The most recent work describing the morphology of high latitude E-region ionization was presented by Baron [1974], Hunsucker [1975], and Vondrak and Baron [1976] - all utilizing the Chatanika incoherent scatter radar as the primary sensor. The Chatanika radar, fully described by Leadabrand, et al. [1972], is a four megawatt incoherent scatter system operating on 1290 MHz utilizing a fully steerable 27m diameter parabolic dish antenna. It is maintained and operated by SRI International and is located 33 km northeast of the University of Alaska, Fairbanks. (Geographic latitude =  $65^{\circ}6'N$ , geographic longitude =  $147^{\circ}27'W$ , geomagnetic latitude =  $65^{\circ}45'N$ , geomagnetic longitude =  $105^{\circ}W$ , geomagnetic inclination =  $76.5^{\circ}$ ,  $L = 5.71$ ). Several groups of investigators, including the Geophysical Institute, use this radar to investigate high latitude thermospheric phenomena.

Figure 1 illustrates the behavior of the relatively undisturbed E-layer observed by the Chatanika radar and the term  $\Delta N/\Delta t$  defines an "impulse characterization" used by Hunsucker [1975] to describe the effects of various types of disturbances on the high-latitude E-region. The ionization peak at 1330 UT is a rather typical sporadic-E ( $E_s$ ) occurrence for the time of day and year. The symbol  $N(120)$  indicates that the electron number density obtained was at a height of 120 km. Throughout this paper the standard URSL nomenclature for ionograms will be used, as specified by Piggott and Rawer [1972], and Piggott [1975].

Figure 2 shows a simultaneous Chatanika radar electron density profile and a College ionosonde ionogram (the College ionosonde is located 33 km southwest of the Chatanika radar). This is an example of the E-layer electron density values observed during a very active auroral disturbance, [Hunsucker and Owren, 1962]. The sporadic-E top frequency ( $fE_s$ ) observed by the College ionosonde was  $\approx 11.2$  MHz and the corresponding electron density obtained by Chatanika radar (bottom plot of Figure 2) shows an E-layer electron density of  $1.4 \times 10^6 \text{ el/cm}^3$  [Hunsucker, 1965].

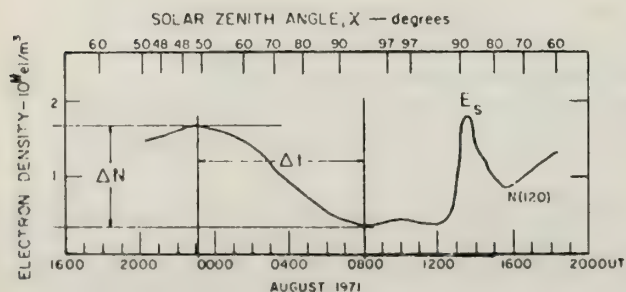


FIGURE 1

Example of regular (undisturbed) E-layer behavior observed by the Chatanika radar and definition of  $\Delta N/\Delta t$  parameters for "impulse characterization" [Hunsucker, 1975].

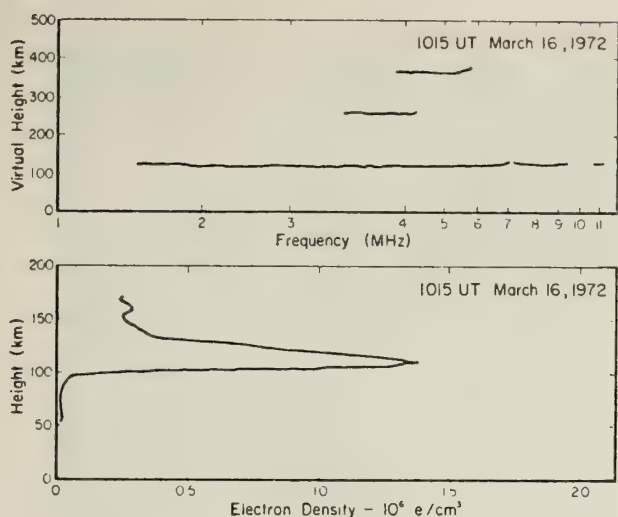


FIGURE 2

Extreme examples of the response of the high latitude E-region to particle precipitation events. This figure shows the response of the E-region to auroral electron precipitation (simultaneous incoherent scatter and ionosonde data). System resolution was 10 km. [Bates and Hunsucker, 1974].

The relationship between visual aurora and electron density distribution in the E-region is illustrated in Figures 3 and 4. The solid line plot is the Chatanika radar electron density, the insert on the right is a tracing of the College ionogram, and the inset at the lower left is the Chatanika all-sky camera photograph (ASC). Figure 3, showing the simultaneous data for 0706 UT (2106 Alaska Standard Time), illustrates the "background level" of E-region ionization of  $\approx 0.2 \times 10^6$  electrons/cm<sup>3</sup>, the right insert a  $fE_s$  of 7.5 MHz and the ASC indicating a discrete auroral arc north of Chatanika. At 0800 UT (2200 AST) in Figure 4, the ASC shows a discrete auroral arc in the beam of the Chatanika radar resulting in an E-region electron density of  $1.2 \times 10^6$  el/cm<sup>3</sup> and an  $fE_s$  of 11.2 MHz. Using the earth-curvature-limited secant factor of 5.0 results in an E-layer maximum-usable-frequency (MUF) of 56 MHz. Maximum observed frequencies (MOF's) of 46 MHz have been recorded on transpolar paths instrumented with forward sounding systems [Hunsucker and Bates, 1969].

The rapid variation in time of E-region ionization during an auroral event is further illustrated in Figure 5, for April 2, 1973. The electron density maximum at 0950 UT was enhanced by the passage through the E-region of an auroral-infrasonic-wave (AIW) as described in detail by Wilson, et al. [1974]. The auroral arc associated with the AIW crossed the Chatanika radar beam at 0951 UT and the E-region electron density increased to  $2.8 \times 10^6$  el/cm<sup>3</sup> at  $h = 100$  km. This electron density corresponds to a plasma frequency of 14.9 MHz and if one uses the earth-curvature-limited secant factor of 5.0, a MUF of 74.45 MHz is obtained for several seconds! The only other experimental justification of this is the qualitative finding that when auroral arcs are present between Anchorage and Fairbanks, Alaska, one occasionally observes bursts of VHF TV transmission from channels in the other city.

Evidence that the auroral E-layer sometimes supports propagation at VHF frequencies during the "auroral season" over transpolar paths as long as 5000 km is presented in Figure 6. These data were obtained on synchronized forward sounding circuits to College, Alaska, from Thule,



March 2, 1973  
0706UT

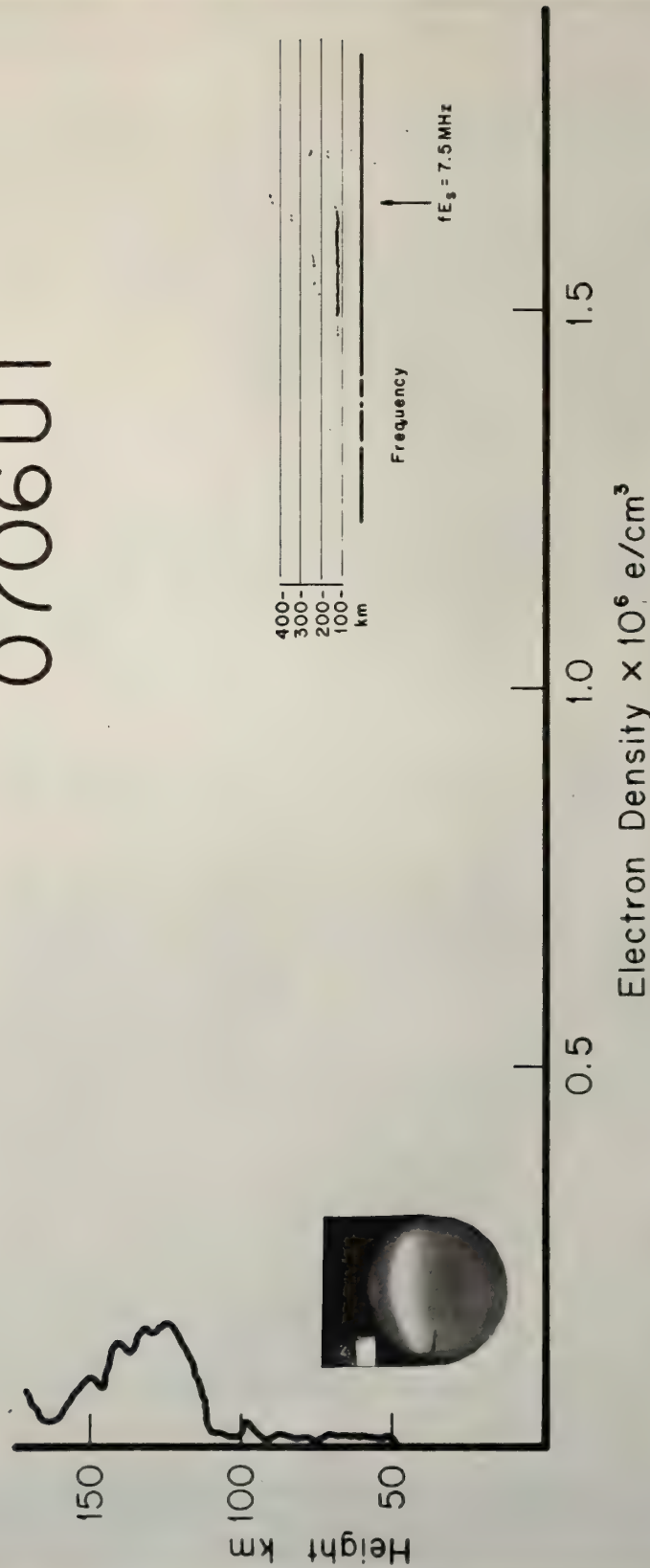


FIGURE 3

Simultaneous Chatanika radar electron density profile, Chatanika all-sky camera picture, and tracing of College ionosonde for 0706 UT March 2, 1973.  
[Hunsucker, 1975]

# 0800UT

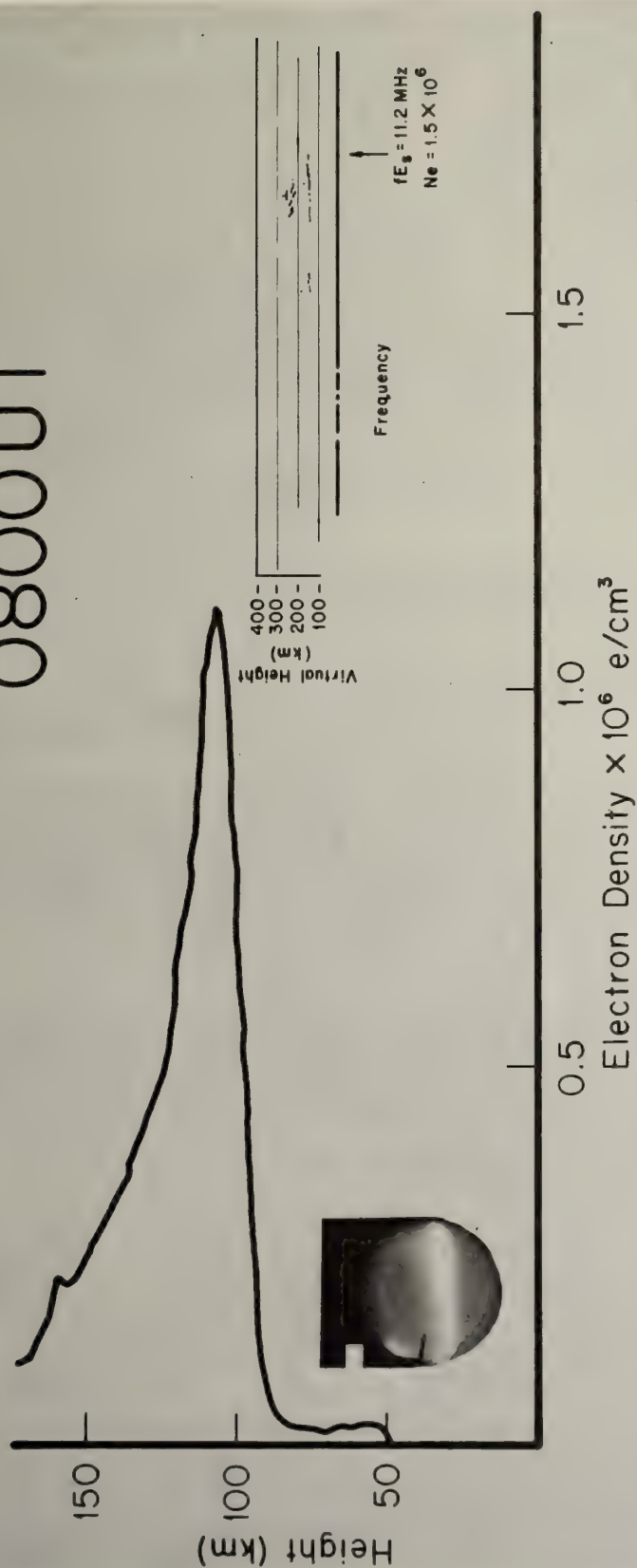


FIGURE 4

Simultaneous Chatanika radar electron density profile, Chatanika all-sky camera picture, and tracing of College ionosonde for 0800 UT March 2, 1973.  
 [Hunsucker, 1975]

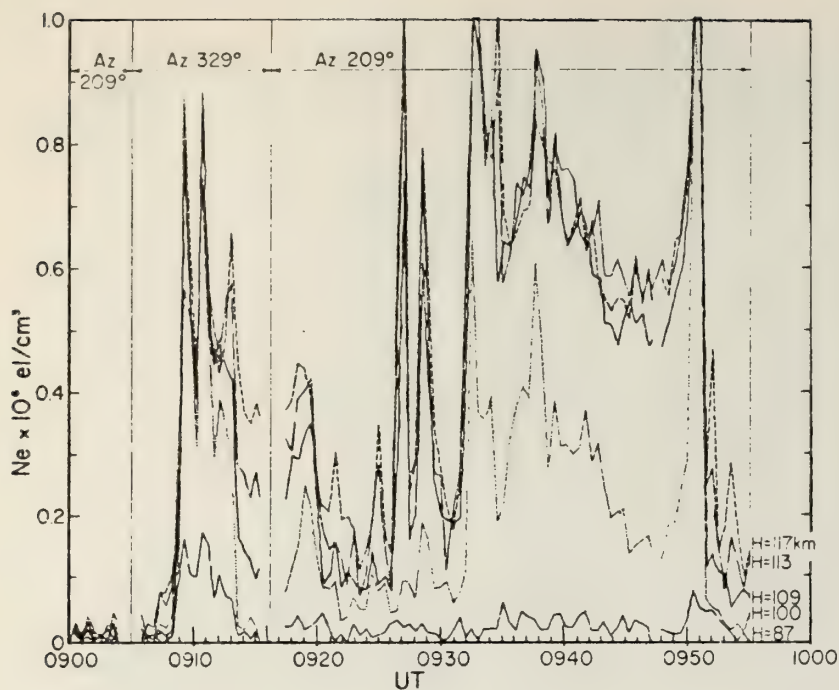
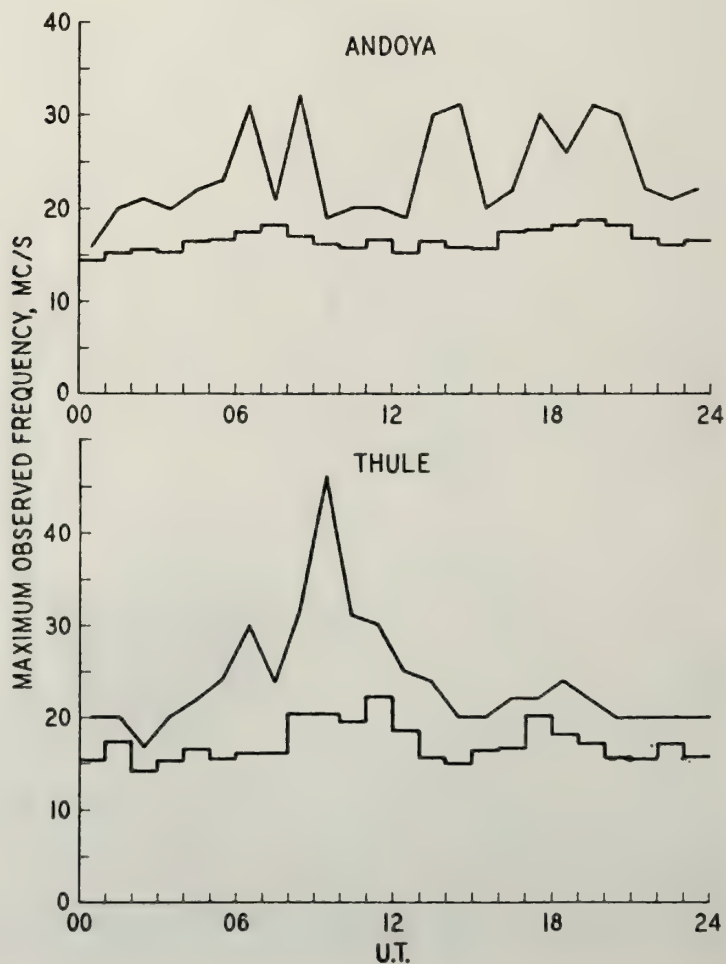


FIGURE 5

E-region electron density time variation during the AIW event of April 2, 1973, obtained with the Chatanika radar. [Hunsucker, 1975].

FIGURE 6

Sporadic-E MOF's from Andoya and Thule for the period from November 27, 1963 to February 12, 1964. The histograms give the hourly averages and the upper curves denote the highest MOF for the hour. [Hunsucker and Bates, 1969.]





Greenland, and Andoya, Norway, during winter-night sunspot minimum conditions. Sporadic-E ( $E_s$ ) modes had an average occurrence of over 50%, with  $E_s$  MUF's as high as 46 MHz with typical values of 18 MHz on the Thule - College path, and MUF's as high as 32 MHz with typical values of 16 MHz on the Andoya - College circuit.

Table 1 summarizes the effects of various disturbance phenomena on the auroral zone E-region. The first horizontal row lists the general behavior of the daily solar variation for comparison with the disturbance phenomena. The disturbances are listed with their dynamic nature generally increasing downward. The solar eclipse disturbance is only a factor of 3 greater than the daily solar variation, the AIW and auroral particle effects are roughly equal, but increase by a factor of  $10^3$  over the previous two, and the magnetic storm phenomenon is the most dynamic of all.

Figure 7 shows a map of various transpolar paths for which HF-VHF forward-sounding data have been obtained for different levels of auroral activity [Hunsucker and Bates, 1969]. The statistical auroral zone drawn on the map illustrates approximately where such paths might be influenced by auroral E-region ionization. (Also, see Jull, et al. [1962] for similar studies.)

## B. Non-great-Circle (NGC) propagation

An interesting and potentially important HF propagation effect is the NGC phenomenon, which is particularly important at high latitudes. The normal HF propagation path follows approximately a great circle from transmitter to receiver, but for signals propagating in the middle- to high-latitude region this is not always the case. Whale [1959] summarized the results of a number of investigations and attributed the NGC modes to ionospheric irregularities in the auroral zone. Bates, Albee, and Hunsucker [1966] found that the incidence of NGC signals on paths near the auroral zone were produced by E- and F-region auroral irregularities in a manner similar to auroral backscatter.

TABLE 1. Summary of the effects of various disturbances on the auroral zone E-layer.

Date	Disturbance phenomenon	Representative values						Comments
		$N_{max}$ (el cm <sup>-3</sup> )	Height (km)	$\frac{N_{max}}{N_{min}}$	$\Delta N$ (el cm <sup>-3</sup> )	$\Delta t$ (sec)	$\frac{\Delta N}{\Delta t}$ (cm <sup>-3</sup> sec <sup>-1</sup> )	
	Daily solar variation (EUV)	$1.5 \times 10^5$	~120	5.0	$1.2 \times 10^5$	$3.24 \times 10^4$	3.7	May-Aug
June 30, 1954	Solar eclipse	$6.0 \times 10^4$	~100	1.7	$2.4 \times 10^4$	$3.0 \times 10^3$	8.0	Baker Lake, CA.
July 20, 1963	(EUV)	$2.8 \times 10^4$	~110	4.0	$7.3 \times 10^4$	$3.9 \times 10^3$	18.7	Anchorage, AK.
July 10, 1972		$1.2 \times 10^4$	110	1.7	$5.0 \times 10^3$	$3.3 \times 10^3$	1.5	Chatanika, AK.
				Avg = 2.5			Avg = 9.4	
Apr. 2, 1973	"Bow shock wave" associated with auroral infrasonic wave (AIW)	$2.8 \times 10^6$	102	14.0	$2.6 \times 10^6$	$3.3 \times 10^2$	$7.9 \times 10^3$	College, AK.
Nov. 21, 1971	Auroral particle	$\sim 1.5 \times 10^6$	108	30.0	$1.45 \times 10^6$	$1.8 \times 10^2$	$8.1 \times 10^3$	Chatanika, AK
Feb. 24, 1972	precipitation	$7.5 \times 10^5$	112	9.3	$6.8 \times 10^5$	$1.2 \times 10^2$	$5.7 \times 10^3$	
Mar. 16, 1972		$\sim 1.5 \times 10^6$	109	10.0	$1.35 \times 10^6$	$2.4 \times 10^2$	$5.6 \times 10^3$	
Mar. 2, 1973		$1.2 \times 10^6$	100	24.0	$1.14 \times 10^6$	$1.8 \times 10^2$	$6.3 \times 10^3$	
				Avg = 18.3			Avg = $6.4 \times 10^3$	
Aug. 6, 1972 (0809 UT)	Magnetic storm	$4.43 \times 10^6$	101	3.9	$3.29 \times 10^6$	$4.8 \times 10^1$	$6.8 \times 10^4$	Chatanika, AK (largest $N_e$ observed by Chatanika radar)

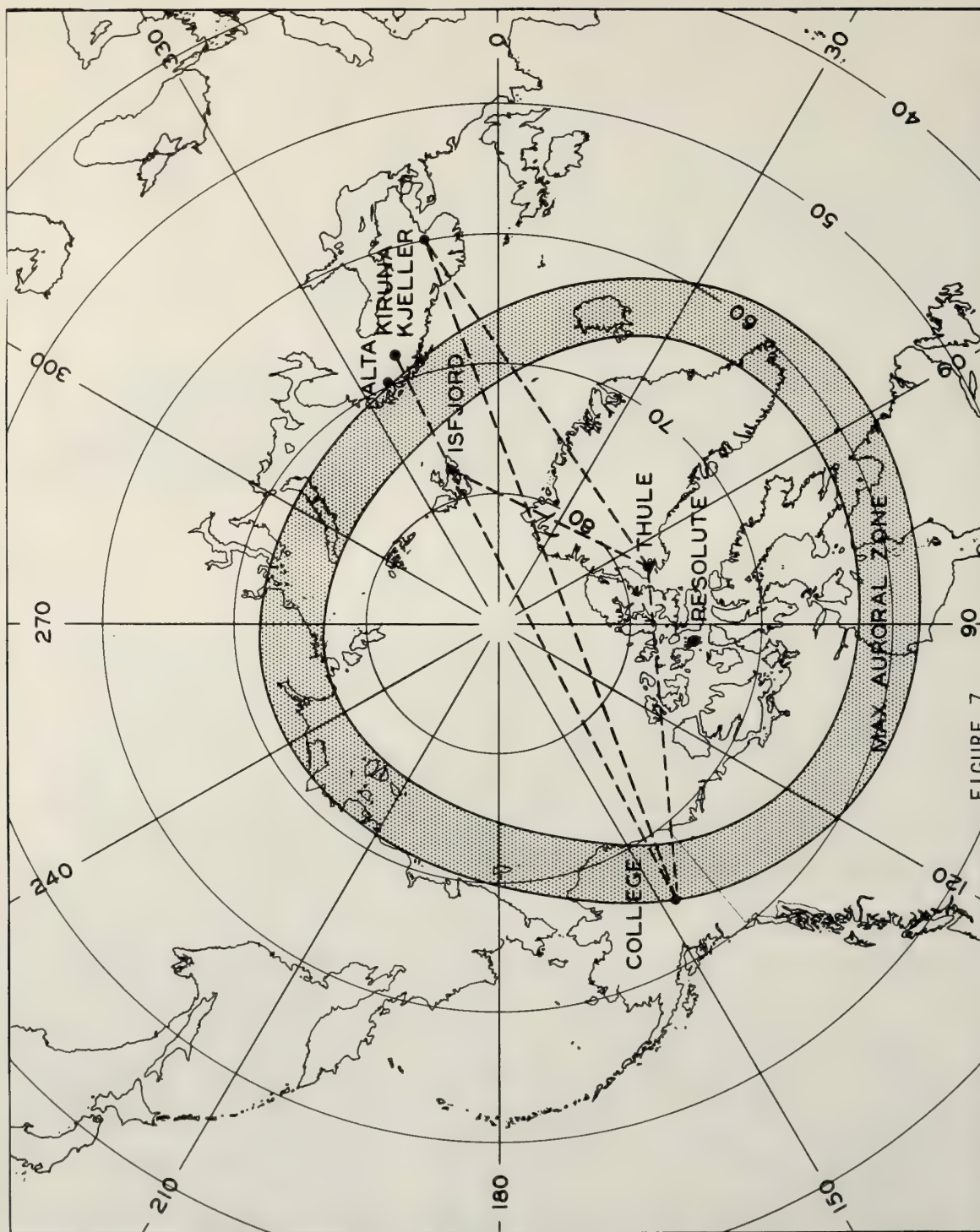


FIGURE 7

Approximate great-circle paths in relation to the statistical auroral zone for various HF circuits. [Hunsucker and Bates, 1969].



On long middle- and high-latitude paths, NGC signals due to auroral sidescatter could be of importance to communications, which could at times be maintained by utilizing the auroral NGC modes. A number of instances were found for which the direct-path signal recorded at College diminished markedly in amplitude, while the NGC signal increased during a disturbance.

In certain cases, however, NGC modes would probably be a liability rather than an asset. Strong NGC signals could considerably degrade direction-finding accuracy. High-speed pulse transmissions might be impossible to receive correctly if the signals were of comparable amplitude and the NGC travel time exceeded the direct-path time by approximately the pulse width or more. The majority of the excess delay times recorded at College on paths from Okinawa, Johnston Island, California, New Jersey, Greenland, and Norway were 2 to 6 msec; Figure 9 illustrates two NGC signals delayed approximately 4 and 11 msec after the direct-path signal (propagation time approximately 12 msec). Particularly on the longer paths, the MOF of the NGC mode was frequently as great as that of the direct-path signal. Very spread signals, such as the one illustrated in Figure 8, were often observed on polar propagation paths monitored at College.

### C. F-Region Ionization

F<sub>1</sub>-layer effects. During recent sunspot minima, the high-latitude F<sub>2</sub> layer was not appreciably more ionized during the summer than the F<sub>1</sub> layer. Normal F-layer forward ionograms are illustrated in Figures 10a and 10b, whereas Figures 10c and 10d show how the F<sub>1</sub> layer modified the shape of the traces. Forward ionograms such as these are readily explainable as being produced by propagation through a thick, dense, lower layer before being reflected by the upper layer.

Figure 11 illustrates an ionogram somewhat similar in shape to that shown in Figure 10, except that only a single strong mode was recorded with a higher than normal delay. Bates and Albee [1966] suggested that it was produced by ducted propagation between the F<sub>1</sub> and the F<sub>2</sub> layers.

During the 1964 solar minimum, the F<sub>1</sub> layer frequently carried the MUF during midmorning summer hours on the path from Palo Alto to College, but relatively fewer instances were noted during a short test in 1965. The probable reason is the increasing dominance of the F<sub>2</sub> layer over the F<sub>1</sub> on the College to Palo Alto path as the bottom of the sunspot cycle passed. As sunspot maximum approaches, the F<sub>1</sub> layer becomes less and less important for long-distance propagation purposes, except possibly for polar cap paths for the reason noted above.

An improved F<sub>1</sub>-layer prediction system for latitudes greater than 40°N and path lengths of 2000-3400 km has been devised by Petrie and Stevens [1965]. They present contour maps of zero distance F<sub>1</sub>-MUF and M factors as a function of sunspot number and local time for April,



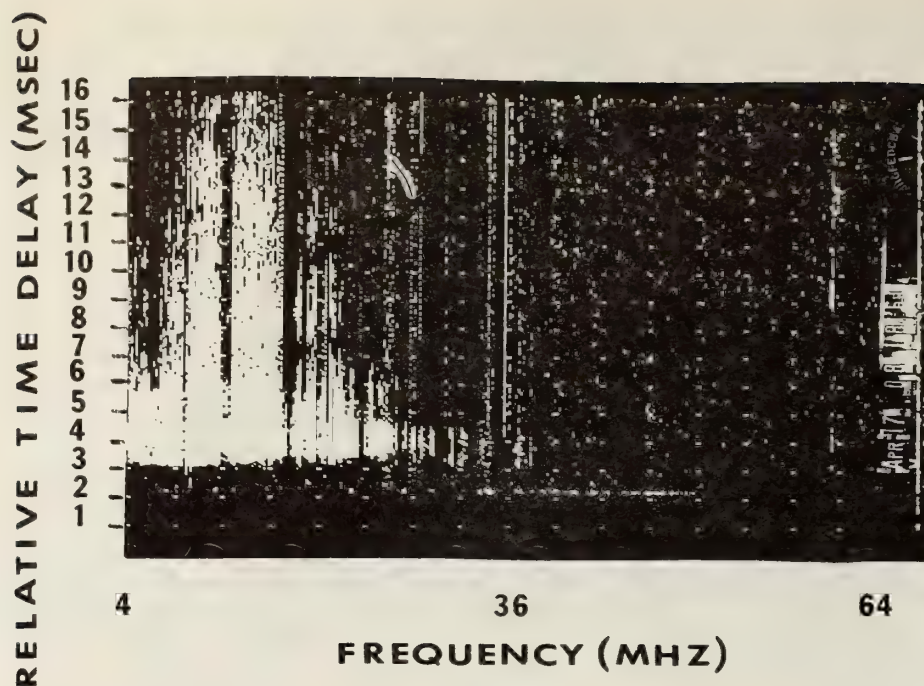


FIGURE 8

A case of extreme spreading of the signal on the step frequency  
Thule, Greenland, to College, Alaska, sounding circuit.  
[Hunsucker and Bates, 1969].

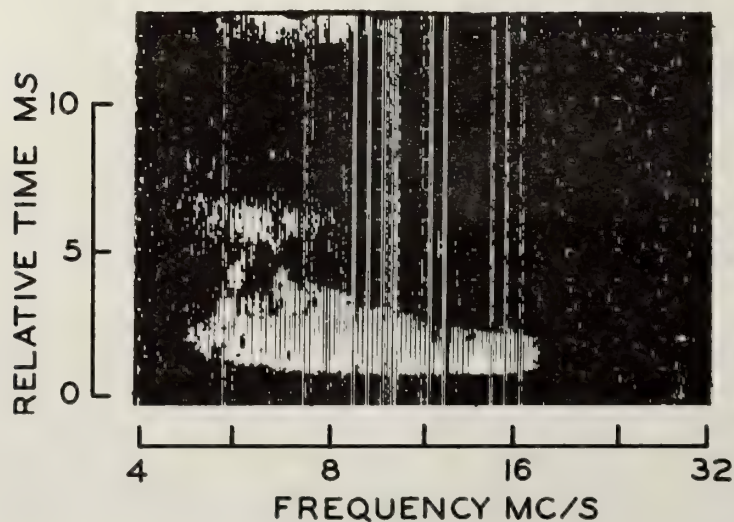


FIGURE 9

An extreme case of off-path propagation on the Palo Alto to College  
path. The trace at the top of the record was delayed by 11 msec for  
a total propagation time of 23 msec. [Hunsucker and Bates, 1969].

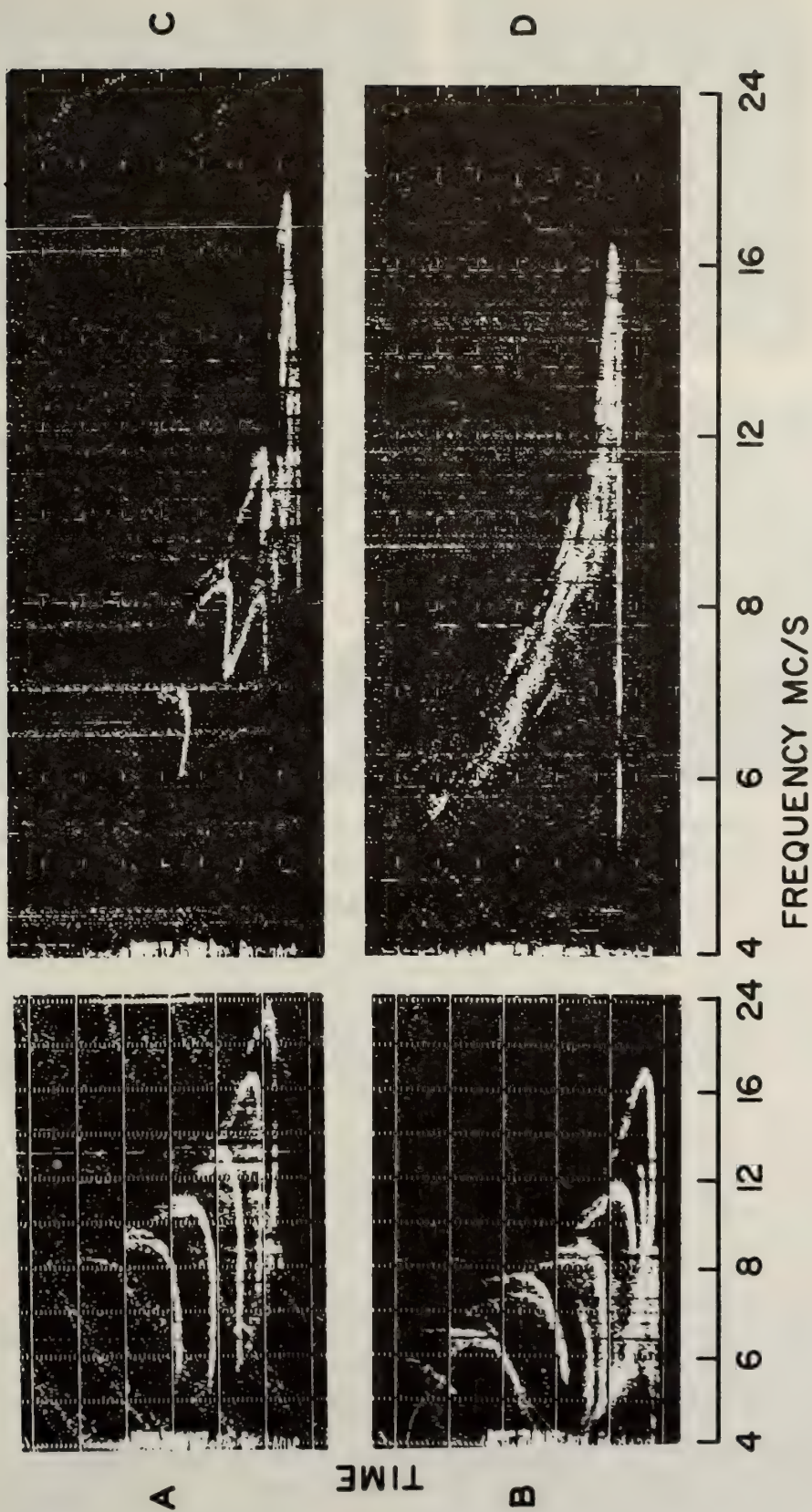


FIGURE 10

Forward oblique ionograms recorded on the Palo Alto to College path. (A and B) Recorded during October; (C and D) recorded during July, 1965. Time marks are 1 msec apart. [Hunsucker and Bates, 1969].



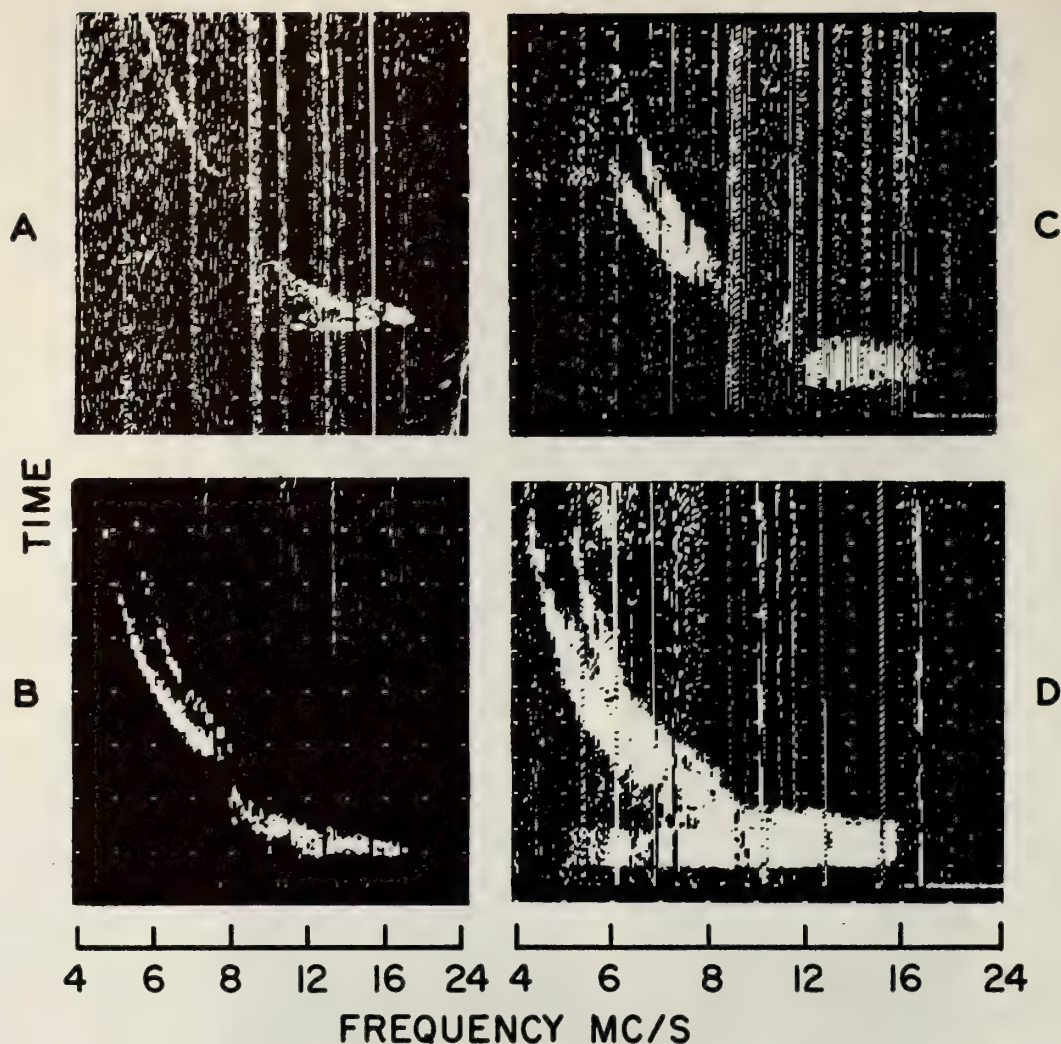


FIGURE 11

Records showing highly delayed signals. Records A and B were obtained on the Palo Alto to College path with 1000- $\mu$ sec pulses on the Fort Monmouth to College and the Palo Alto to College paths, respectively. [Hunsucker and Bates, 1969].

June, August, and October. For optimum results this information should be included in high-latitude HF propagation predictions, particularly during periods of low sunspot activity.

Figure 12 shows typical quiet-day signal strength variations on several of the transpolar paths indicated in Figure 7. An interesting point is that high signal levels were recorded for approximately 13 hours of the day (54%) when  $M(3000)F_2$  was less than the operating frequency.

The determination of propagation modes from forward sounding data is illustrated in Figure 13 for winter conditions.



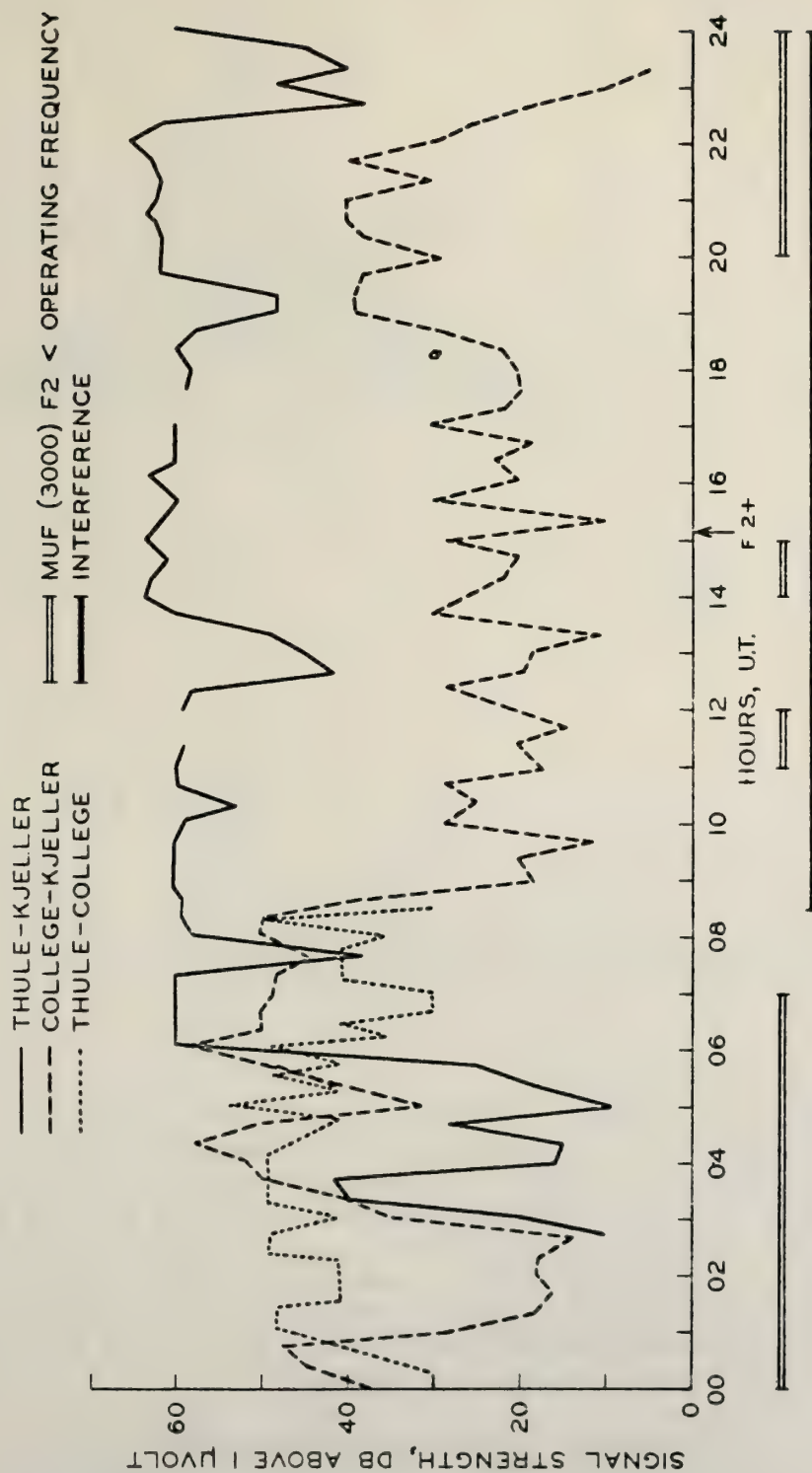


FIGURE 12

Quiet-day variations of 18 MHz signal strength for various polar HF paths. June, 1961, Sunspot Number = 77.

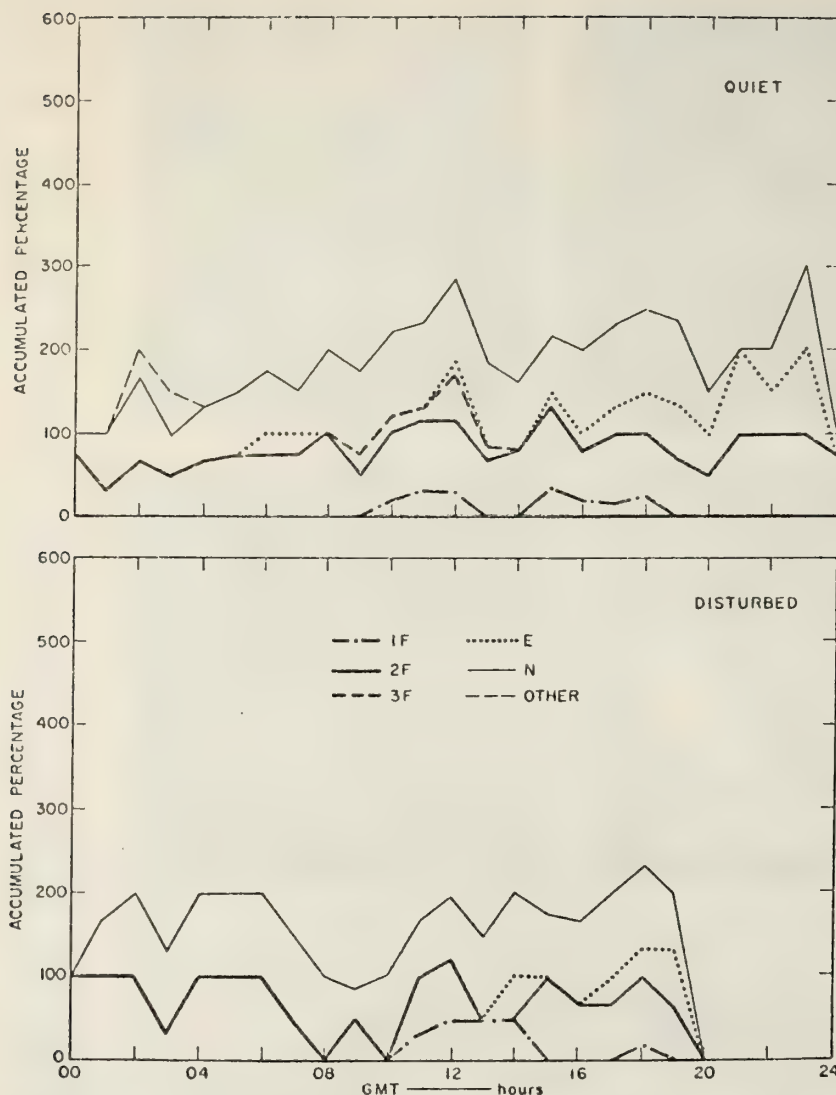


FIGURE 13

January configuration of modes Andoya, Norway to College, Alaska, for quiet and disturbed magnetic conditions. [Hunsucker and Bates, 1969].

#### D. Fading on high-latitude paths.

As noted in the section on NGC propagation, communications might be maintained at times by using auroral sidescatter modes, but this technique presents some problems. Auroral backscatter echoes exhibit a fast fading rate (tens to hundreds of cycles per second), (Auterman [1962] and Koch and Petrie [1962]) and auroral sidescatter echoes probably will also do so. Such a high fading rate, or "auroral-flutter" as it is sometimes called, may prevent high-speed communications that use pulse widths of the order of milliseconds.

Lomax [1967] studied several polar paths to determine the signal spectrum; Figure 14 illustrates typical power spectra he obtained on temperate and high-latitude paths. (See also Hadaller and Schrader [1965]).

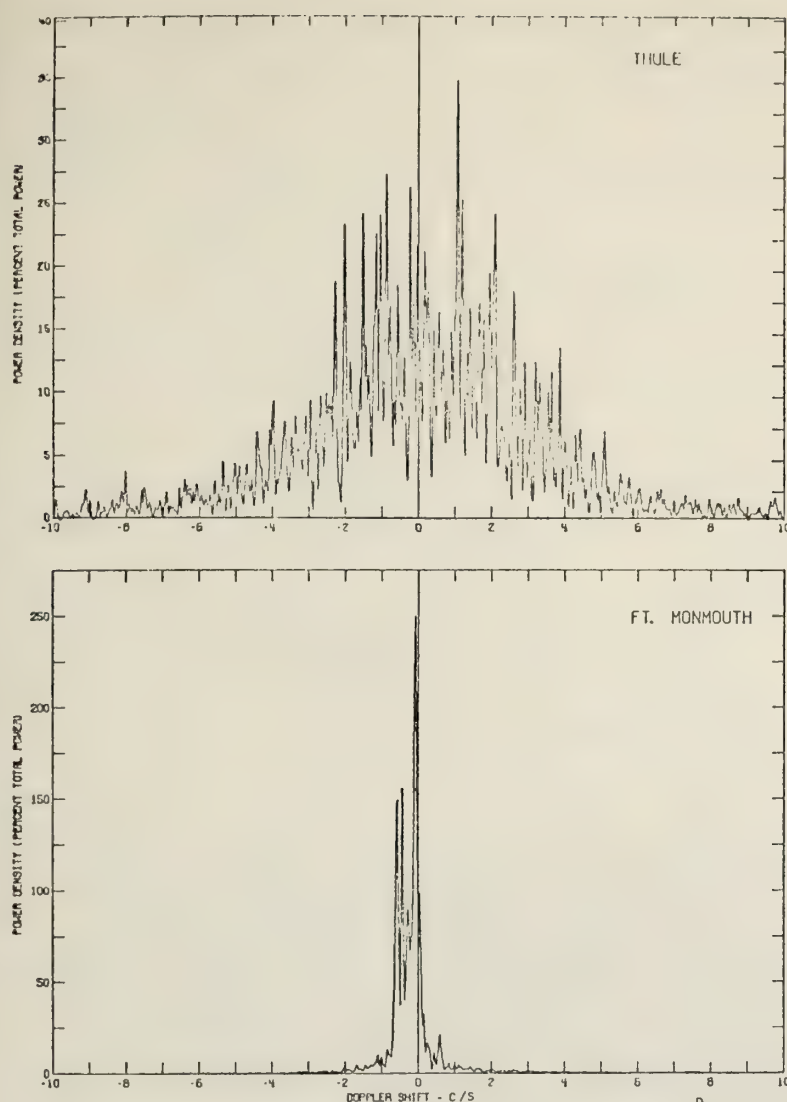


FIGURE 14

Typical power spectra transmissions from Thule and Fort Monmouth monitored at Palo Alto. [Hunsucker and Bates, 1969].

### E. Spread-F

Figure 15 illustrates with simultaneous incoherent scatter radar and vertical ionosonde data another quite common high latitude phenomenon, "spread-F". High latitude spread-F appears frequently on an ionogram as a group of separate vertical-incidence traces which approach a range of critical frequencies, as illustrated by the ionogram in Figure 15. A basic problem in using spread ionograms for HF propagation purposes has been to obtain the true F-layer critical frequency from the spread ionogram. A considerable amount of past work on this question has culminated in the recommendation that  $f_{oF}$  be scaled from the inner edge of the spread ordinary trace [URSI, 1972; Piggott, 1975].

Using the Chatanika radar data, we have found that in each of the few clear cases examined, the density at the F-layer maximum is indeed given by the critical frequency of the inner edge of the spread trace. The system constant of the radar is known precisely enough that the



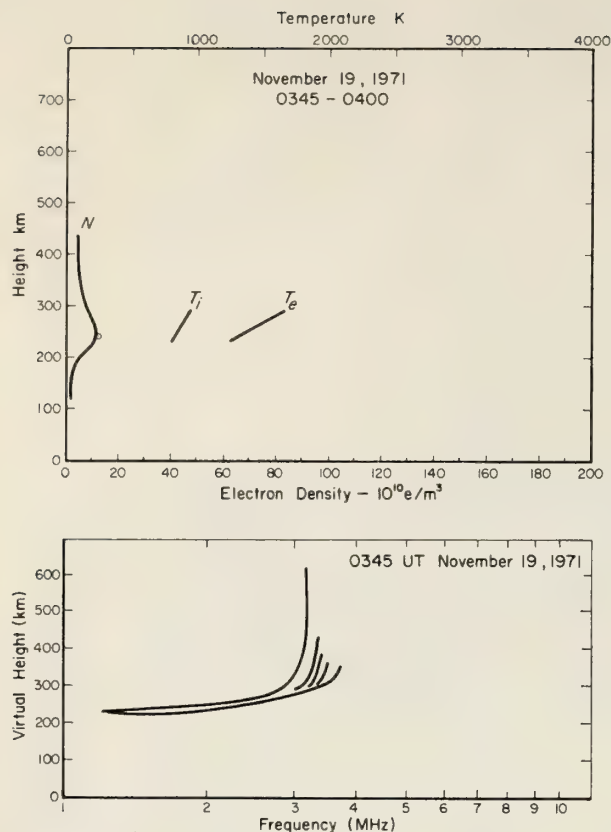


FIGURE 15

Simultaneous incoherent scatter radar electron density profile and vertical ionogram for spread-F conditions. System resolution was 50 km. [Bates and Hunsucker, 1974].

incoherent scatter radar value is meaningful. A representative example of our results are presented; the points on the scatter radar profile in Figure 15 represents the critical density scaled from the inner edge of the ordinary trace in the simultaneously recorded ionogram at College. Thus, our results, admittedly derived from a limited data sample, suggest that the commonly accepted method of scaling  $f_oF$  from spread ionosonde traces is indeed correct.

### III. SUMMARY

Table 2 (on the following page) summarizes the most important high-latitude E- and F-layer anomalies and their effects on HF to VHF propagation. These anomalies may be regarded as second order perturbations on the normal behavior of the E and F-layers as described by currently used midlatitude HF propagation prediction programs.

Acknowledgements: This work was supported by NOAA Contract Number 03-6-022-35-102; Atmospheric Science Section, National Science Foundation, NSF Grant ATM77-21460; and State of Alaska research funds. The typing and layout of this paper was accomplished by Ms. Pat Brooks.

TABLE 2

Propagation Phenomenon	Sallient Effects	Remarks
NGC modes	(a) Deviations from great-circle path up to 90°. (b) MUF sometimes carried by NCG mode. (c) Garbling of high data-rate transmissions (fast fading rate or interference between modes).	At times may be the only signal propagating on a particular path.
Sporadic E	High MOF on circuits traversing the auroral zone.	Primarily important during winter-night period. Sometimes associated with the aurora.
F <sub>1</sub> layer	Frequently the propagation mode during the summer morning.	Most pronounced during sunspot minimum conditions.
Fading	Fading rate ~20 Hz. Mean fading correlation bandwidth ~4 kHz.	Associated with auroral zone electron density irregularities. Should peak in equinoctial months. Maximum during morning hours.

## REFERENCES

- Agy, Vaughn (1957): Proceedings of the polar atmosphere symposium, part 2, Ionospheric section, J. Atmospheric Terrest. Phys., special supplement, 129-134.
- Auterman, J. L. (1962): Fading correlation bandwidth and short-term frequency stability measurements on a high-frequency transauroral path, NBS Tech. Note 165.
- Baron, M. J. (1974): Electron densities within aurorae and other auroral E-region characteristics, Radio Science, 9, 341-348.
- Bartholomew, Richard R. (1966): Results of a high-latitude HF backscatter study, Scientific Rept. 2 (SRI Project 5538), Stanford Research Institute, Menlo Park, CA.
- Bates, H. F., and P. R. Albee (1966): On the strong influence of the F<sub>1</sub> layer on medium to high latitude HF propagation, Sci. Rept. UAG-R175, Geophysical Inst., Univ. of Alaska, College, Alaska.
- Bates, H. F., P. R. Albee, and R. D. Hunsucker (1966): On the relationship of the aurora to non-great-circle HF propagation, J. Geophys. Res., 71(5), 1413-1420.
- Bates, H. F., and R. D. Hunsucker (1974): Quiet and disturbed electron density profiles in the auroral zone ionosphere, Radio Science, 9, 455-467.

- Folkestad, Kristen (Ed.) (1968): Ionospheric Radio Communications, Plenum Press, New York.
- Gassmann, G. J. (Ed.) (1963): The effect of disturbances of solar origin on communications, AGARDograph 59, Pergamon Press, New York.
- Hadaller, H. L., and D. H. Schrader (1965): Polar propagation during short-wave fadeouts, J. Geophys. Res., 70(5), 1265-1269.
- Hunsucker, R. D. (1965): On the determination of electron density within discrete auroral forms in the E region, J. Geophys. Res., 70, 3791-3792.
- Hunsucker, Robert D. (1967): HF propagation effects at high latitudes, Or - look north, young ham!, QST, LI, no. 2, 16-19 and 132.
- Hunsucker, Robert D. (1975): Chatanika radar investigation of high-latitude E-region ionization structure and dynamics, Radio Science, 10, 3, 277-288.
- Hunsucker, R. D., and H. F. Bates (1969): Survey of polar and auroral region effects on HF propagation, Radio Science, 4, 347-365.
- Hunsucker, R. D., and L. Owren (1962): Auroral sporadic-E ionization, J. Res. Nat. Bur. Stand., Sect. D, 66D, 581-592.
- Jull, F. W., D. J. Doyle, and J. P. Murray (1962): Frequency sounding techniques for HF communications over auroral paths, Proc. IRE, 50(7), 1676-1682.
- Kirby, R. C. and C. G. Little (1960): Conference on Arctic Communication, J. Res. NBS Radio Propagation, 64D(1), 73-80.
- Koch, J. W., and H. E. Petrie (1962): Fading characteristics observed on a high-frequency auroral radio path, J. Res. NBS Radio Propagation, 66D(2), 159-166.
- Landmark, F. (Ed.) (1964): Arctic Communications, Pergamon Press, New York.
- Leadabrand, R. L., M. J. Baron, J. Petriceks, and H. F. Bates (1972): Chatanika, Alaska, auroral-zone incoherent-scatter facility, Radio Science, 7, 747-756.
- Lomax, John B. (1967): High-frequency propagation dispersion, Stanford Research Institute, Menlo Park, Calif.
- Penndorf, R., and S. C. Coroniti (1959): Propagation of HF and VHF in the arctic region, IRE Trans. Commun. Systems, 7(6), 121-125.



- Petrie, L. E., and E. E. Stevens (1965): An  $F_1$  layer MUF prediction system for northern latitudes, IEEE Trans. Antenna Propagation, 13(4), 542-546.
- Piggot, W. R. (1975): High-latitude supplement to the URSI handbook of ionogram interpretation and reduction, URSI Report UAG-50, World Data Center A, NOAA, Boulder, CO 80302.
- URSI (1972): URSI handbook of ionogram interpretation and reduction, 2nd ed., edited by W. R. Piggot and K. Rawer, Rep. UAG-23, pp. 143-147, World Data Center A. NOAA, Boulder, CO 80302.
- Vondrak, R. R., and M. J. Baron (1976): Radar measurements of the latitudinal variation of auroral ionization, Radio Science, 11, 11, 939-946.
- Watt, T. M. (1973): Incoherent scatter observations of the ionosphere over Chatanika, Alaska, J. Geophys. Res., 78, 16, 2092-3006.
- Whale, H. A. (1959): The effects of ionospheric irregularities and the auroral zone on the bearings of shortwave radio signals, J. Atmospheric Terrest. Phys., 13(3/4), 258-270.
- Wilson, C. R., R. D. Hunsucker, and G. J. Romick (1976): An auroral infrasonic substorm investigation using Chatanika radar and other geophysical sensors, Planet. Space Sci., Vol. 24, 1155-1175.

## C. REPORT OF THE MID- AND LOW-LATITUDE E AND F REGION WORKING GROUP

A WORKING GROUP REPORT by C. Rush, Chairman

### 1. INTRODUCTION

The E and F regions of the ionosphere at the low and middle latitudes have been the subject of intense study for over 50 years. The influence of the structure of the E and F region on the propagation of radio waves that are reflected from the ionosphere and the predictability of this influence has been addressed for over 35 years. While a level of sophistication has emerged in the ability to understand, specify, and predict the low and mid-latitude E and F regions, this same level of sophistication enables ionospheric physicists, propagation system operators and telecommunication policymakers to demand an increasing knowledge of these regions of the ionosphere. This results primarily from the fact that the level of current understanding with regard to the formation and variation of the ionosphere permits those individuals who must deal with the ionosphere to ask the right questions to ensure further understanding and exploitation of the ionosphere.

In the following sections, the current status of our ability to predict the low and mid-latitude E and F regions is briefly summarized.

### 2. E-REGION VARIABILITY

The E-region extends from 90-130 km and consists of both the normal and sporadic E-layers. The normal E-layer has a strong solar zenith angle dependence with a maximum density near noon and a seasonal maximum during the summer. Typically, the E-layer critical frequency,  $f_oE$ , is about 3 MHz and the height of the maximum electron density occurs at about 110 km. At night the value of  $f_oE$  drops to 0.4 - 0.6 MHz (Wakai, 1971). The E-layer electron density closely parallels the solar cycle: E-layer maximum densities occur at solar maximum. The  $f_oE$  varies about 30 percent over the solar cycle at a given location (CCIR, Report 725, 1978).

The E-layer is produced by photoionization and a complex process involving reactions and recombinations between the ionized and neutral atmosphere. Ionization results from ultraviolet ( $E_{UV}$ ) and X-ray emissions acting on the neutral atmosphere. The  $E_{UV}$  and X-ray fluxes are measurable quantities. Unfortunately, limited information is known about neutral atmospheric densities at 110 km. Therefore, detailed mathematical prediction models appear to be of limited use for operational purposes. Current prediction techniques combine limited mathematical modeling with statistical techniques. These techniques are applicable to both long-term and short-term predictions.

## 2.1 Long-Term Variability

A number of methods have been proposed to forecast long-term (monthly median) values of foE. As stated earlier, foE varies with both season and solar cycle. Presently, long-term monthly median predictions depend on statistical data from a variety of worldwide locations. Two prediction methods currently used are those developed by Muggleton (1975) and by Leftin (1976). The Muggleton model (1975) is based upon published ionospheric data from the years 1944-1973 observed at 55 ionospheric stations. The model provides values of foE as a function of solar activity (10.7 cm flux), solar declination, and geographic latitude. Nighttime foE values are based upon observations of Wakai (1971). The model developed by Leftin (1976) yields worldwide predictions of foE based upon sunspot number, month, universal time, geographic latitude, and geographic longitude. Nighttime foE values are also based upon the work of Wakai (1971).

Both of these models rely on forecasting solar-related parameters (10.7 cm flux or sunspot number) and unless improvements are made in solar forecasting, improvements in long-term E-layer parameter predictions will be limited. In addition, further work is required to validate and to extend current knowledge of the nighttime E-layer structure.

## 2.2 Short-Term Variability

Of all the ionospheric layers, the normal E-layer is the most regular. The critical frequency of the E layer fluctuates only slightly from day to day with a range of only  $\pm 6$  percent (Rawer, 1952). Enhanced X-ray flux from solar flares can increase foE from five to ten percent. Therefore, short-term predictions of foE require forecasting both changes of the neutral atmosphere coupled with variations of the  $E_{UV}$  and X-ray fluxes. Since neutral density changes and flux variation forecasting is very limited, most short-term predictions depend on forecaster experience.

A recent paper (Ivanov-Kholodny and Nusinov, 1979) describes a new technique for specifying the present state of the E-layer based upon measured variation of  $E_{UV}$  and X-ray fluxes versus mean solar flux values. This technique offers hope in that if one can obtain a good description of current E-layer structure, one can better forecast short-term behavior based on morphology and/or experience.

## 3.0 F REGION VARIABILITY

The F region extends from about 130 km into the topside ionosphere (i.e., the ionosphere above the height of the maximum electron density denoted  $LmF2$ ). The F region conventionally is discussed in terms of two distinct regions--an F1 region which is strongly dominated by solar control and an F2 region which displays a behavior that is quite different from that characteristic of solar control.

The F1 region follows a Chapman-type of behavior and usually is confined



to the heights between 130 and 210 km. As distinct from the F1 region, the F2 region does not exist during the nighttime hours. The F2 region displays significant departures from a Chapman model with dynamical processes resulting from electrodynamic drifts, ambipolar diffusion, and neutral air winds, playing a major role in the control of the ionization distribution.

The behavior of the critical frequency of the F2 region is often talked about in terms of "anomalies." The use of the word anomalies dates back to the time at which ionospheric physicists held the Chapman model to be valid for all ionospheric regions and departures from expected Chapman behavior were termed anomalies. Thus, the names diurnal anomaly (referring to the fact that foF2 often reaches its peak value not at noon but typically between 13 to 15 hours local time), seasonal anomaly (referring to the fact that the noon-time values of foF2 in winter tend to exceed the corresponding summer values) and equatorial anomaly (referring to the marked departure from solar zenith angle dependence of foF2 within  $\pm 20$ - $30^\circ$  of the geomagnetic equator) have become standard nomenclature in ionospheric technology.

The height of the F-2 region, hmF2, varies significantly during the course of the day. At low latitudes, hmF2 is typically 50-150 km higher during the day than at night. At mid-latitudes, on the other hand, the situation is reversed with nighttime values of hmF2 being 50-100 km higher than daytime values.

Of the normal ionospheric parameters, values of foF2 and hmF2 tend to show the greatest variability on any time (and spatial) scale. It is for this reason that much of the emphasis in predicting ionospheric structure is geared toward these parameters. Values of foF2 and hmF2 form the basis for many models of the ionosphere ranging from empirical models of the electron density structure to propagation system simulation models.

### 3.1 Long-Term Variability

Many forecast centers continue to provide long-term and short-term predictions of F region parameters. Values of foF1 can be obtained on a long-term basis by employing the methods described by Du Charmé et al. (1973) or by Rosich and Jones (1973). Long-term values of foF2 generally are determined by employing the coefficients described in the CCIR Atlas of Ionospheric Characteristics (CCIR Report 340-3). The coefficients used to determine foF2 result from the work of Jones et al. (1969) and Jones and Obitts (1970). The coefficients developed by Jones et al. (1969) define the diurnal and geographical representation of foF2 and are based on a linear relationship with solar activity. The coefficients developed by Jones and Obitts define the diurnal, geographical, seasonal, and solar cycle variations of foF2 with a second degree polynomial being used to represent the solar cycle variation.

The height of the F2 region maximum electron density generally is determined indirectly by use of relationships between hmF2 and the M(3000)F2 factor. Long-term predictions of M(3000)F2 are available through coefficients contained within propagation prediction programs such as described by Barghausen et al. (1969).

### 3.2 Short-Term Variability

Variations of 15-20 percent from (monthly) median foF2 over intervals of a day or less are known to occur during quiet times as well as during magnetic storms. These may be superimposed on slower upwards or downwards drifts in values over several days. It is desirable to predict all these variations. The need for predicting the day-to-day variability has been pointed out in detail in CCIR Report 727 (1978). Many methods have been developed to update and specify foF2 (and hmF2) using the long-term predictions as a point of departure.

A method developed by SRI, for example, (Hatfield, 1979) uses the CCIR coefficients as a departure point for a computer code that performs point-to-point raytracing incorporating improvements for high latitude F region structure and layer tilts. At the Air Force Global Weather Central (Tascione et al., 1979), a four-dimensional model, again using the CCIR coefficients in an initial representation, is updated with near-real time ionosonde and total electron content (TEC) observations. Forecasting centers in France (Lassudrie-Duchesne, 1979), Japan (Maeda, 1979) and Australia (McNamara, 1979; Wilkinson, 1979) perform predictions of foF2 and M(3000)F2 on a weekly or monthly basis with daily update.

Several possibilities and many requirements for improving predictions of F2 region parameters were discussed in the workshop. Work in the Soviet Union (Zevakina and Lavrova, 1979), for example, reveals a relationship between the direction of the interplanetary magnetic field (IMF) and the sign of the foF2 deviation from its median variation. Since the IMF is monitorable, this affords a true improvement for short-term prediction. However, few such really new possibilities seem to be available. Rather, it appears that significant improvements in predictions will come about mainly by increasing the data base, and by improving statistical, temporal, and spatial representation. The CCIR coefficients are based on data observed primarily between 1954 and 1964. Much more information could be brought into the representation, for example, by using the foF2 analysis undertaken by Paul (1976). Nevertheless, geographic coverage is still quite uneven, with much more data available from the northern than southern hemisphere.

### 4.0 IRREGULARITIES IN THE E REGION - SPORADIC E ( $E_s$ )

Sporadic E is the name given to dense blobs of ionization in the E region that possess horizontal dimensions between 100 and 1000 km, and vertical dimensions between 1 and 10 km (Kerblay and Nosova, 1979). The physical causes of Sporadic E are still being investigated. Mid-latitude Sporadic E is usually attributed to wind shear phenomenon and low latitude Sporadic E is attributed primarily to plasma instabilities in the equatorial electrojet. The importance of  $E_s$  for communication purposes is based on the evidence that Sporadic E layer can act as a reflecting layer for HF and VHF emissions (CCIR Report 259-3, 1978; Miya et al., 1978). Sporadic E can affect long-distance communication by creating interference to wanted signals. For example, the effects of Sporadic E can result in a much larger coverage area for an UHF



transmitting station than predicted with the consequence being interference to similar stations located within this area.

The CCIR is at present evaluating the percent of time of  $E_s$  occurrence in the equatorial regions. Work in progress indicates that the occurrence of  $E_s$  is much higher at low latitudes than at mid-latitudes, rising to 90 percent of time near the magnetic equator for 3 MHz vertical incidence frequency.

CCIR Report 259-3 (1978) provides a set of curves indicative of percent of time that  $E_s$  is present for different hours and seasons for mid-latitudes section of America, Europe, and Asia. As an indication of the order of magnitude, these percentages vary from 0.01 percent up to 30 percent of time, depending upon the maximum frequency. Giraldez (1979) has presented a method of prediction for  $E_s$  critical frequency, which depends on latitude, season, hour and solar activity, for the daytime hours. This method has been tested with observational data from North and South America, Europe, Africa, Asia, and Oceanic, and the error of predicted value is within ten percent in all cases. Kerblay and Nosova (1979) provide an indication of other prediction methods for  $E_s$  critical frequency developed in USSR.

CCIR Report 259-3 (1978) also provides a method for calculating signal strength from  $E_s$  reflections, as a function of critical  $E_s$  frequency, distance from transmitter and antenna characteristics.

## 5.0 IRREGULARITIES IN THE F-LAYER

### 5.1 Low Latitudes

The onset of low latitude F-region plasma irregularities begins one hour or so after sunset with scale sizes ranging from tens of kilometers down to three meters. Occurrence frequency and seasonal characteristics at different longitudes are well documented for both range and frequency type spread F (Rastogi, 1978). Range type spread F lends itself well to short-term prediction since it occurs after sunset only when the bottomside F layer is rising or has risen above a critical altitude of about 350 km (Balsley et al., 1972). Also, if foF2 shows a fairly sharp decline of 1 to 2 MHz after 1800 to 1850 LT, this implies that the F-layer is moving up rapidly which precedes the occurrence of spread F by about an hour.

The theory of the range type spread F is not well understood but is thought to be associated with the formation of large scale depleted regions of electron density originating at bottomside F region heights after sunset. The growth of the large scale irregularities is due to a Raleigh-Taylor type instability mechanism (Haerendel, 1973) or possibly to an  $\vec{E} \times \vec{B}$  drift instability process. Once formed, the depleted region buoyantly rises through the ambient ionosphere creating smaller scale irregularities in its wake. A great deal more study is needed to understand the growth mechanics and the conditions under which spread F can be sustained. The rising plasma bubbles themselves may act as HF waveguide for trans-equatorial propagation. The occurrence frequency of equatorial spread F generally maximizes during equinoctal periods



and during solar cycle maximum years.

## 5.2 Mid-Latitudes

Irregularities in the F region electron density can result in changes in the amplitude, phase, and angle of arrival of radio waves transmitted through or reflected by the F region. Spread-F which is observed on ionograms is caused by irregularities in the F-region electron density.

The dimensions of small-scale ionospheric irregularities are found to range from several kilometers to hundreds of meters in the L-shell range of approximately 2-4 (Basu et al., 1976), and the overall dimensions of F-layer irregularity patches are found to be on the order of hundreds of kilometers.

Singleton (1979) has modified the global irregularity model developed by Fremouw and Rino (1973) by taking into account longitudinal and magnetic activity effects. For the mid-latitudes, Singleton's model assumes a value, for the axial motion of field-aligned irregularities ranging from 10 to 5 with the lower value at the higher latitudes.

There is a lack of experimental data on the occurrence, scale size and spatial extent of mid-latitude F-layer irregularities as a function of time of day, season, sunspot number and magnetic activity. The data are required for the validation of the theoretical models.

## 6. THEORETICAL STUDIES

The usefulness of theoretical models to predict electron density distribution in the low- and mid-latitude ionospheric E and F regions is limited by the ability to predict the parameters which enter into the relevant equations being solved; i.e., neutral atmospheric constituents (Straus and Hickman, 1979); neutral and charged particle temperatures, neutral wind, and electric fields, and ionizing sources such as solar  $E_{UV}$  radiation and energetic particles (Spjeldvik and Lyons, 1979). With the exception of plasma irregularities, the physical processes affecting the low- and mid-latitude E and F regions are basically well understood. The fact that the input parameters are difficult to predict accurately, particularly for short-term applications, implies that theoretical modeling itself may be rather limited.

The limitations mentioned above do not mean that theoretical models are not useful. They are used, advantageously, to improve our understanding of the coupling processes which occur between 1) electron and ion temperatures and densities, 2) neutral dynamics and charged particle density (Roble, 1975), 3) neutral dynamics and electric fields (Heelis et al., 1974). They can also be used to determine how sensitive the electron density structure is to variations in the input parameters. Where observational data are lacking, either in space or time, theoretical models can be used to estimate electron densities.

Areas for long-term research must, therefore, include improving our basic

understanding of the input parameters and how they respond to changing solar and geophysical conditions. In particular, extensive studies of the ionospheric electric fields need to be undertaken since this is one of the most important parameters in effecting low latitude electron density distribution and is one of the least well understood. Also needed is a better understanding of the processes involved in the local generation of electric fields at low latitudes such as the F region dynamo (Rishbeth, 1971) and plasma instability generated electric fields (Anderson and Haerendel, 1979). At mid-latitudes, primary input parameters include the neutral atmosphere and neutral winds and the development of a general circulation model of the thermosphere is crucial to understanding the response of the neutral atmosphere to a variety of energy sources.

## 7. DISTURBANCES OF THE IONOSPHERE AT MID- AND LOW-LATITUDES

Major disturbances in the mid-latitude F region are associated with and follow large disturbances at higher latitudes in the auroral zone. At the higher mid-latitudes, values of foF2 usually decrease and virtual heights increase during storms; the variation with local time, and with time after storm onset, is complex and variable. At lower latitudes, storm effects may result in either higher or lower values of foF2, probably dependent on season, while in the equatorial region there is usually an increase in foF2 together with a redistribution of the ionization forming the equatorial anomaly.

At mid-latitudes, characteristic variations of foF2 have been shown to be associated with the form of the variation of geomagnetic field (Zevakina and Lavrova, 1979). However, the morphology and development of mid-latitude disturbances need further study since the descriptions now available are inadequate for communications prediction purposes.

The daytime normal E region shows little change during disturbance; the F1 region is also little changed but its importance may greatly increase due to the reduction in F2 region ionization concentration. At night, an intermediate layer ( $h' \approx 140$  km) is observed at mid-latitudes; its occurrence depends on geomagnetic activity. High values of foEs occur at mid-latitudes due to gravity wave propagation following high latitude disturbances (Giraldez, 1979). Further studies are required of storm-related sporadic E ionization and of the nighttime E stratifications.

## 8. CONCLUSIONS

As a result of the activities of the Mid- and Low-Latitude E and F Region Working Group throughout the workshop, the following efforts were thought to be necessary in order to secure an improved ability to predict and specify the low and mid-latitude E and F regions.

It is Recommended that:

1. . . . studies and efforts be undertaken to predict and specify the

electron density in the lower ionosphere (E and F1 regions) during the nighttime hours.

2. . . . emphasis be given to the development of realistic, operationally useful methods to predict Sporadic E and its effects on the performance of HF propagation systems.
3. . . . the operation of the network of ionospheric sounding stations be continued and upgraded with modern instrumentation. Further emphasis should be placed on obtaining ionospheric information in those regions of the globe inaccessible to the present routine observations.
4. . . . efforts be conducted to improve the long-term predictions of the F2 region parameters, particularly in the equatorial region and in the southern hemisphere.
5. . . . studies be conducted to assess the degree to which electric fields in the equatorial region which are of importance to the dynamics of the low latitude E and F regions are related to electric fields observed at high latitudes.
6. . . . studies be continued to further develop theoretical and empirical models of the ionosphere for application to the short-term prediction of ionospheric characteristics.
7. . . . observations be made of the following geophysical parameters and that studies be conducted relating short-term changes in F2 region characteristics to these parameters: interplanetary magnetic field, solar wind velocity, and magnetic activity.
8. . . . since short-term changes in the maximum height of the F2 regions are important in considering the performance of a wide variety of telecommunications systems, that high frequency measurement methods be adopted which assure adequate temporal and frequency continuity.
9. . . . experimental data on the location, scale-size, and spatial extent of mid-latitude F-layer irregularities be collected and studied as a function of time of day, season, sunspot and magnetic activity.

## 9. REFERENCES

- Anderson, D. N. and G. Haerendel (1979): The motion of depleted plasma regions in the equatorial ionosphere, J. Geophys. Res. (in press).
- Balsley, B. B., G. Haerendel and R. A. Grunwald (1972): Equatorial spread F: recent observations and a new interpretation, J. Geophys. Res., **77**, 5625.
- Barghausen, A. F., J. W. Finney, L. L. Proctor, and L. D. Schultz (1969): Predicting long-term operational parameters of high-frequency sky-wave telecommunications systems, Tech. Rpt. ERL 110-ITS-28, ESSA Research Laboratories, Institute for Telecommunication Sciences.



- Basu, Sunanda, S. Basu, and B. K. Khan (1976): Model of equatorial scintillations from in-situ measurements, Radio Science, 10, 821.
- CCIR, Report 259-4 (1978): VHF propagation by regular layers, sporadic-E and other anomalous ionization, Geneva.
- CCIR, Report 340-3 (1978): CCIR atlas of ionospheric characteristics, Geneva.
- CCIR, Report 725 (1978): Ionospheric properties, Geneva.
- CCIR, Report 727 (1978): Short-term prediction of operational parameters for ionospheric radio communications, Geneva.
- Du Charme, E. D., L. E. Petrie, and R. Eyfrig (1973): A method for predicting the F1-layer critical frequency based on the Zurich smoothed sunspot number, Radio Science, 8, 837.
- Fremouw, E. J., and C. L. Rino (1973): An empirical model for average F-layer scintillation at VHF/UHF, Radio Science, 8, 213.
- Giraldez, A. E. (April 1979): Daytime sporadic-E blanketing frequency prediction, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #15).
- Haerendel, G. (1973): Theory of equatorial spread F, Report, Max Planck Institut fur Physik und Astrophysik, Munich.
- Hatfield, V. E. (April 1979): HF communications predictions 1978 (An economical up-to-date computer code, AMBCOM), Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #68).
- Heelis, R. A., P. C. Kendall, R. J. Moffett, D. W. Windle and H. Rishbeth (1974): Electrical coupling of the E and F regions and its effect on F region drifts and winds, Planet. & Space Sci., 22, 743.
- Ivanov-Kholodny, G. S., and A. A. Nusinov (April 1979): Critical frequency and height of maximum forecast of mid-latitude E-layer, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #147).
- Jones, W. B., and D. L. Obitts (1970): Global representation of annual and solar cycle variation of foF2 monthly median 1954-1958, Telecommunications Research Report OT/ITS RR3, Boulder, CO.
- Jones, W. B., R. O. Graham, and M. Leftin (1969): Advances in ionospheric mapping by numerical methods, Tech. Report ERL 107-ITS 75, ESSA Research Laboratories, Institute for Telecommunication Sciences.
- Kerblay, T. S., and G. N. Nosova (April 1979): Statistical prediction of Es-layer parameters and echo-signal characteristics, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #159).
- Lassudrie-Duchesne, P., A. M. Bourdila, and H. Sizun (April 1979): The French short-term radio propagation predictions in the decameter band,

Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #7).

Leftin, M. (1976): Numerical representation of monthly median critical frequencies for the regular E-region (foE), Office of Telecommunication Rpt. 76-88, U.S. Gov't. Printing Office, Washington, D. C.

Maeda, R. (April 1979): Radio propagation prediction services in Japan, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #10).

McNamara, L. F. (April 1979): The use of ionospheric indices to make real- and near real-time forecasts of foF2 around Australia, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #116).

Miya, K., K. Shimizu and T. Kojima (1978): Oblique incidence sporadic-E propagation and its ionospheric attenuation, Radio Science, 13, 559.

Muggeleton, L. M. (1975): A method of predicting foE at anytime and place, Telecommunication J., 42, 413.

Paul, A. K. (1978): Temporal and spatial distribution of the spectral components of foF2, J. Atmos. Phys., 40, 135.

Rastogi, R. G. (1978): On the equatorial spread F, Proc. Indian Acad. Sci., 87, 115.

Rawer, K. (1952): Calculation of sky-wave field strength, Wireless Engineer, 29, 287.

Rishbeth, H. (1971): Polarization fields produced by winds in the equatorial F region, Planet. & Space Sci., 19, 357.

Roble, R. G. (1975): The calculated and observed diurnal variation of the ionosphere over Millstone Hill, Planet. & Space Sci., 28, 1017.

Rosich, R. K., and W. B. Jones (1973): The numerical representation of the critical frequency of the F1 region of the ionosphere, Office of Telecommunications Report OT 73-22, Institute for Telecommunication Sciences.

Singleton, D. G. (April 1979): Predicting transionospheric propagation conditions, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, Co, (Paper #60).

Spjeldvik, W. N., and L. R. Lyons (April 1979): On the predictability of radiation belt electron precipitation into the earth's atmosphere following magnetic storms, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #51).

Straus, J. M., and D. R. Hickman (April 1979): Predictability of upper-atmospheric density and composition, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #19).

- Tascione, T. F., T. W. Flattery, V. G. Patterson, J. A. Secan, and J. W. Taylor, Jr. (April 1979): Ionospheric modeling at Air Force Global Weather Central, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #106).
- Wakai, N. (1971): Study on the night-time E region and its effect on the radio wave propagation, J. Radio Res. Labs. (Japan), 18, 245.
- Wilkinson, P. J. (April 1979): Prediction limits for foF2, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #118).
- Zevakina, R. A., and E. V. Lavrova, On the possibility to predict variations in the F2 region parameters as a function of the IMF direction, Proceedings of the Solar-Terrestrial Predictions Workshop, Boulder, CO, (Paper #88).



## D. D-REGION PREDICTIONS

A WORKING GROUP REPORT prepared by: E. V. Thrane, Chairman, Members: D. K. Chakrabarty, S. D. Deshpande, R. H. Doherty, J. B. Gregory, J. K. Hargreaves, J. Lastovicka, P. Morris, W. R. Piggott, J. B. Reagan, W. A. Schlueter, E. Swanson and W. Swider.

### 1. SUMMARY

Present knowledge of D-region phenomena is briefly reviewed and the status of current methods of predicting their effects on radio propagation considered. ELF, VLF and LF navigational and timing systems depend on the stability of the lower part of the D layer where these waves are reflected, whereas MF and HF waves are absorbed as they penetrate the region, in most cases mainly in the upper part of the layer. Some gaps in basic knowledge are identified. Possible methods of improving predictions, warnings and real-time operations are considered with particular stress on those which can be implemented in the near future. This report is supported by two invited review papers by Swider and Larsen, which follow this report and numerous short papers in Volume III of this proceeding. Action needed is identified in eight recommendations.

### 2. RECOMMENDATIONS

The Working Group on D-Region Predictions is acutely aware of the often pragmatic approaches required to provide timely predictions to users. Practical schemes are limited by telemetry, technical competence of users, and physical predictive facilities as well as scientific knowledge. Severe limitations exist in all areas. Recommendations are of two types: (1) those to improve the utility of present knowledge and (2) those to improve fundamental knowledge on which future substantial improvements depend.

1. The Working Group recommends the development of telemetry for timely dissemination of disturbance information to users. Even one bit to indicate the presence or absence of an event would be useful. Two methods are worth considering. A unique, non-interfering modulation could be added to world wide Omega signals, or to HF signals from WWV or elsewhere to indicate the presence of an event. The former offers the advantage of being continuously easily receivable on a global basis, even during total HF blackout, and would be appropriate, as Omega users would have much to gain.

2. The Working Group recommends the development of automatic devices indicating appropriate user response to geophysical events.

3. The Working Group recommends that the basic data used to produce the CCIR maps of absorption morphology be recomputed with the Dyson and Bennett correction included, and the resultant maps be submitted to C.C.I.R.

4. The Working Group draws attention to the lack of reliable basic observations of the absorption in the upper part of the D-region and recommends the development, construction and deployment of cheap automatic pulse absorption equipment to provide an early solution to this problem.

5. The Working Group recommends the application of event detectors capable of rapid detection and classification of events such as SID's, PCA's, etc. in real-time on a 24 hour basis, and development of semiempirical SID warnings useful for prediction purposes.

6. The Working Group recommends that new stations be set up to monitor the lower part of the D layer in magnetically low latitudes, below  $\pm 40^\circ$ , since there are no data at present to predict the propagation of OMEGA and other VLF signals in this zone.

7. The Working Group considers that future forecasting in real-time of polar zone radio propagation will be best attained using satellite techniques, to identify the position and intensity of the perturbing phenomenon, and thus provide global mapping of the energetic particle inputs into the D-region. The spatial resolution should be sufficient to identify the location of the dynamic auroral oval, the irradiated polar caps during SPE's and the perturbed midlatitude during and following magnetic storms and substorms.

8. The Working Group recommends that both theoretical studies and experimental investigations, which include satellite techniques, be made into the aeronomy and dynamics of the mesosphere, as in the forthcoming Middle Atmosphere Program, in order to develop a D-region model useable for prediction purposes.

### 3. INTRODUCTION

#### 3.1 The D-Region

The ionospheric D-region is normally defined as the height region between 50 and 95 km above the earth's surface. Compared to the rest of the ionosphere, it is characterized by small ionization densities, and large collision frequencies of electrons and ions with neutral molecules. During normal daytime conditions the concentration of free electrons  $N_e$  increases nearly exponentially from about  $1 \text{ cm}^{-3}$  at 50 km to  $10^4 \text{ cm}^{-3}$  at 90 km, whereas the electron neutral collision frequency  $\nu$  decreases from  $10^8 \text{ s}^{-1}$  to about  $10^5 \text{ s}^{-1}$  in the same height range. The electron densities in the D-region are large enough to cause reflection of extremely low frequency (ELF), very low frequency (VLF), and low frequency (LF) radio waves, and because of the large collision frequencies they cause significant absorption of medium and high frequency radio waves that penetrate the D-region. This absorption depends upon  $\int N_e \nu ds$  along the ray path. During disturbances excess ionization increases the absorbing properties of the layer and changes the reflection height of the long waves.

During normal and moderately disturbed conditions, the product  $N_e \nu$  is largest above 80-85 km. On the other hand, VLF waves are reflected below



80-85 km. Under normal and moderately disturbed conditions it is therefore appropriate to distinguish between the upper D-region, of importance to MF and HF communication, and the lower D-region, of importance to VLF and LF propagation. During severe disturbances, the lower part can also affect MF and HF waves. Table 1 summarizes the characteristics of the disturbances that affect the D-region.

### 3.2 Aim of Study

Working Group C4 had 40 members representing most groups throughout the world which were engaged in active research on the lower ionosphere. Unfortunately users were not well represented.

Two review papers (Larsen preprint 125, and Swider preprint 78) were invited on subjects of importance to the group.

There are two main classes of users for D-region predictions: system planners and designers, and operators. Designers and planners are primarily interested in anticipated steady state conditions and their probability of occurrence. Operators are primarily interested in short-term forecasts or even knowledge of positive detection of events in progress. There are three main categories of users interested in predictions of the D-region:

a) Planners and designers, interested in long-term predictions to determine system coverage and frequency allocation doctrine, and also for the calibration of navigation systems.

b) Users of ELF/VLF and LF navigation and timing systems who depend upon the stability of the region where these waves are reflected for time and position determination.

c) Communicators and broadcasters using the MF and HF bands for long range propagation via the ionosphere who depend upon acceptable signal to noise ratios for their transmissions.

The aim of this working group was to review the present knowledge of the D-region and to discuss to what degree this knowledge can improve prediction schemes for the three groups of users. Note that this group did not deal with the prediction of the solar events that cause D-region disturbances. The group concentrated on the problem of estimating the response of the D-region to a solar disturbance, once this had been (correctly) predicted. In section 4 the report reviews prediction methods in current use. Section 5 deals with the status of D-region modeling, section 6 the needs of the user, and section 7 points out schemes and discusses possibilities for improvements in the future.

## 4. D-REGION PREDICTION METHODS IN CURRENT USE



Table 1. Ionospheric Disturbances.

Disturbance	Approximate Frequency of Occurrence			Typical Duration	Spatial & Temporal Extent	Typical Effects on Long Path at VLF		Typical HF Effect Amplitude	Possible Cause	Comments
	Solar Max.	2/year	Solar Min.			Phase	Amplitude			
Sudden Ionospheric Disturbance (SID)	2/week	2/year		3/4 hr	Sunlit Hemisphere	5-30 cec <sup>3</sup>	-10 to 5 db <sup>5</sup>	Blackout to X < 70°	Enhanced solar X-ray & EUV Flux from Flare	Largest VLF phase effect < 100 cec.
Polar Cap Absorption (PCA)	1/mo	0		3 day	Polar Reg.	40 cec <sup>3</sup>	~0 db <sup>4</sup>	Blackout for illuminated zone. Rarely 10-20 db	Solar Protons 1-100 MeV	Largest VLF phase effect < 100 cec.
Magnetic Storm	26/yr	22/yr		days	Global	10-20 cec <sup>3</sup>	5-10 db <sup>4</sup> fade	n/a <sup>6</sup>		
Pelativistic Electron Participation (REP)	Variation not known; usually present to some degree.			1-5 hr	Sub-Auroral primarily; largest effects during day	10-20 cec <sup>3</sup>	< 2 db fade <sup>4</sup>		Interaction of solar low energy plasma with earth's magnetic field causing energetic electron precipitation.	Larger effects near auroral zones.
Winter Excess Absorption	20/yr <sup>1</sup>	20/yr <sup>1</sup>		days	Winter; usually spatially small	Usually phase advance	5-10 db <sup>4</sup> fade	Normally B Blackout	Precipitation of electrons w/energies of a few hundred Kev.	
Stratospheric Warming	2/yr	2/yr		weeks	Mid-winter to spring; large	Unknown	Unknown	2-4 fold increase in absorption when expressed in db	Changes in Global Circulation pattern.	

1. In Northern Hemisphere.
2. Typical effect on path located for maximum sensitivity.
3. Centicycles. At ten kilohertz 1 cec corresponds to about 1MS in time or 0.1 mile in a hyperbolic system such as Omega.
4. Figures are not well known.
5. Changes of  $\pm 1$  to 2 db common for frequencies near 10kHz; enhancements of 5db expected for frequencies over 14kHz.
6. Perturbations controlled by related phenomena.

Ionospheric predictions fall into two categories I) long-term predictions of monthly median values of radio propagation parameters and II) short-term predictions, often termed forecasts that predict ionospheric conditions days, hours or even minutes in advance. Space does not permit a detailed description of the available methods, but some important factors will be mentioned and references to fuller treatments will be given.

#### 4.1 Long-Term Predictions

Study Group VI of the CCIR is an international body which coordinates studies of the propagation of radio waves via the ionosphere, with the object of improving radio communication. The latest report of this Group (CCIR XIII Plenary Assembly, Geneva 1974 Volume VI) provided descriptions of current long term prediction methods. A report from the Plenary Assembly in Kyoto 1978 is in preparation and will give the latest revisions of the methods. One of these will be discussed in more detail in section 7.

In the present report we are interested in prediction of MF and HF absorption and of LF-VLF phase and amplitude for MF and HF propagation. The most widely used prediction scheme is the "Bluedeck" computer programme, developed at Central Radio Propagation Laboratory in Boulder, Colorado. The absorption estimates in this scheme are based on the work of Lucas & Haydon (1966), Barghausen et al (1969) and Schulz and Gallet (1970). Here the ionospheric absorption is described by semiempirical formulae in terms of frequency, distance, sunspot number, solar zenith angle and geographical position. The CCIR report quoted above (Report 572, Volume VI) refers to improved schemes based on work by George (1971), George and Bradley (1974), and Haydon et al (1976), but these are not in widespread use. The absorption predictions are useful and reasonably accurate for low latitudes and for middle latitudes in summer. The later references quoted above also take into account winter anomaly absorption, and, in a very rough way, high latitude absorption. We shall discuss the limitations of the methods in section 7.

In the CCIR Report 264-3 (1974), field strength curves for the frequency range 150 kHz - 1600 kHz are presented, and Report 265-3 (1974) deals with skywave propagation at frequencies below 150 kHz. Whereas the first of these reports is based upon purely empirical formulae adapted to measurements, the field strength and phase curves in the latter reports are derived from theoretical computations from simple ionospheric models. Of special interest for this working group is the prediction scheme used for the Omega navigation system. Omega predictions are very possibly the most used ionospheric predictions of any type, being in daily use by thousands of navigators. The Omega system has global coverage and uses frequencies in the range 10-13 kHz. The predictions of the phase velocity of the Omega signals have been described by Swanson (1972), Swanson & Brown (1972), Swanson (1977), and Morris and Cha (1974). The method assumes that the phase of a signal may be expressed as a sum of incremental phase shifts over path increments and include both variations in the ground and ionosphere. The method is useful, and essential for the Omega users, but has certain limitations that will be discussed later. Present system accuracy is significantly degraded by prediction inaccuracies.



Several AGARD publications deal with ionospheric predictions and the reader should note: AGARD Conference Proceedings No. 49 (1170) "Ionospheric Forecasting", AGARD CP 173 (1976) "Radio Systems and the Ionosphere" and AGARD CP 238 (1978) "Operational Modelling of the Aerospace Environment".

#### 4.2 Short Term Predictions

Predictions for periods equal to or less than the solar rotation period, 27 days, are called forecasts. Disturbances in progress are described by warnings. A number of centers throughout the world issue warnings and short term forecasts of solar and ionospheric parameters (Davies 1978). Twelve of these are grouped into the International Ursigram and World Days Service (IUWDS) for the exchange of data and cooperation in solar geophysical observations. In the USA the most important forecasting centers are US Air Force Global Weather Central in Omaha and NOAA Space Environment Forecast Center in Boulder. The USSR also issues short time forecasts (Zevaknia et al 1962). The forecasts normally give qualitative statements on the degree of disturbance expected, for example "moderate HF absorption" or "general improvement of MF propagation conditions". Forecasts for the degree of VLF phase disturbance are often issued for PCA's; predictions for SID's have only recently been attempted experimentally (Swanson & Levine 1978). Improvements in recent years in ionospheric forecasting are due mainly to more efficient data acquisition and assessment. Examples are the real time propagation assessment systems "Prophet" and the real time navigation monitor developed by the US Naval Ocean System Center (Rothmuller, 1978), (Swanson & Levine, 1978).

### 5. PRESENT STATUS OF D-REGION MODELLING

One of the main tasks of WG C4 has been to collect the most recent information on the physics and chemistry of the D-region and to discuss the present status of D-region modelling. There are two approaches to D-region modelling. One is best represented by the Cospar International Reference Atmosphere 1972 and the Reference Ionosphere, Rawer, (1978) which provide average height variations of the main atmospheric parameters, such as pressure, density temperature, ionization density profiles, etc. The models are based upon experimental data using interpolation and extrapolation techniques where information is scarce. The Reference Ionosphere is a new and useful tool. However, on a global scale the experimentally determined D-region ionization density profiles are few, and the uncertainties corresponding great. This is particularly true in regions where there is great variability, such as at high latitudes. Recent work by Davis and Berry (1977) has yielded an empirically described model for the D-region ionosphere by attempting a best fit to all measured profiles available in the literature. They have produced a model usable for VLF or LF phase and amplitude predictions for any given path and time. The measured ionospheric profiles used in their study are available in a data base prepared by McNamara (1978).

The second approach is to build physical models based upon first principles. Given the flux of ionizing radiation, and an initial state and



composition of the neutral atmosphere, is it possible to compute the resulting ionization density distributions and their effects upon radio wave propagation? Considerable effort has been made towards this goal, and in the following we shall discuss some of the results. In doing so it is useful to introduce a subdivision based mainly upon the special processes that dominate in particular latitude regions. First, however, a few general remarks will be made.

## 5.1 D-Region Ion Production and Loss Processes

The basic continuity equation governing production and loss of D-region ionization may be written

$$(1 + \lambda) \frac{dN_e}{dt} = q - \psi N_e^2 - \nabla \cdot N_e \vec{v} \quad (1)$$

Here  $N_e$  is the electron density,  $q$  the ion production rate,  $\psi$  an effective loss rate coefficient,  $\lambda$  the ratio of negative ion density to electron density and  $\vec{v}$  velocity of the air.

Because of the small degree of ionization and relatively large neutral air density, the dynamics of the plasma below 95 km will be determined by the fluid forces of the neutral gas. The resulting transport will determine the concentration of certain key minor constituents. The ion production sources during undisturbed conditions are solar UV, X-rays, galactic cosmic rays and possibly weak electron precipitation and associated bremsstrahlung. During disturbances solar x-rays, energetic proton and electron precipitation with associated bremsstrahlung are important. The ionization and absorption cross sections for all the important radiations and atmospheric constituents are well known, so that in principle the ion production rate may be determined if the height variations of the atmospheric species are specified. It is characteristic of the D-region that selective ionization of minor constituents may dominate the ion production. Ionization of NO by solar H-Lyman- $\alpha$  and ionization of the excited  $O_2(^1\Delta_g)$  by UV are important examples. For this reason the transport of minor constituents into or from the D-region may significantly influence the ionization balance. The loss rate  $\psi$  of ionization through recombination is determined by very complex photochemical processes. Above the mesopause (near 85 km), fast dissociative recombination of electrons with molecular ions  $O_2^+$  and  $NO^+$  dominate and  $\psi$  is of the order of  $5 \cdot 10^{-7} \text{ cm}^3 \text{ s}^{-1}$ . Below this height there is a region where complex hydrated positive water ions dominate, and at the lowest heights neutralization of positive and negative ions determine the loss rate. The cluster ions have large recombination rates, and the value of  $\psi$  may increase from  $5 \cdot 10^{-7} \text{ cm}^3 \text{ s}^{-1}$  at 85 km to  $10^{-3} \text{ cm}^3 \text{ s}^{-1}$  at 60 km. [Arnold and Krankowsky (1977), Knapp & Schwartz (1975), Swider & Dean (1975).] It is important to realize that the D-region photochemistry is not yet completely mapped. In addition, the dynamics of the mesosphere and lower thermosphere are still under investigation, in particular, transport effects are not known quantitatively.

### 5.2.1 The Global D-Region

It is possible to distinguish portions of the globe in which different physical processes dominate the D-region. In so doing, it is necessary to

utilize classifications by i) geographic latitude and hemisphere, in respect to the dynamic meteorology and the aeronomy of the D-region, and ii) magnetic latitude or L-shell value, in respect to processes originating in the magnetosphere. The result is an overlapping of regions separately classified, but which shows the superimposed influence of different processes. It is to be noted that any classifications developed for one hemisphere may not be considered as directly applicable to the other. The curves of constant invariant and dip latitudes are much further displaced from the corresponding geographic latitudes, and from each other, in the Southern hemisphere than in the Northern. Again, the dynamics of the stratospheric and tropospheric circulations, which have interactions with mesospheric circulations, differ in the two hemispheres. For example, there are striking differences in the morphology of the increased radiowave absorption in winter ("winter anomaly") in the Northern and Southern hemispheres.

Utilizing this two-fold approach to the global D-region, the following main divisions are recognizable:

- (a) The low-latitude (geographic) D-region, bounded approximately at  $\pm 30^\circ$ .
- (b) The mid-latitude (geographic) D-region, poleward of  $\approx 35^\circ$ . Within this region, dynamical phenomena may be continuous in latitude to the pole, eg. the winter eastward circulation in the mesosphere, whereas aeronomic phenomena may show change, eg. at the polar circle in winter.
- (c) The high latitude (geomagnetic) D-region, poleward (magnetic) of  $60^\circ$  invariant latitude or L-shell value of 4

These classifications may be qualified in respect to certain phenomena, eg. to include geomagnetic latitudes down to  $L = 2$  when considering various aspects of energetic particle precipitation. Nevertheless, the division, quoted form an adequate basis for the discussion below.

### 5.2.2 The Low Latitude D-Region

The D-region at geomagnetic latitudes below  $30^\circ$  is simpler to model than at other latitudes for two reasons. Firstly, the ionosphere is not disturbed by energetic particle precipitation, except in the vicinity of the South Atlantic anomaly. Secondly, the dynamics of the low latitude mesosphere, though not well known, do not appear to be influenced by the larger scale wave motions of the stratosphere and troposphere, such as occur in the mid-latitude D-region. The low latitude D-region is most likely characterized by three main ionizing components; a variable flux of solar X-rays dominating the ionization rate above 85 - 95 km., a steady flux of solar H-Lyman-ionizing nitric oxide in the height range 65 - 85 km, and galactic cosmic rays providing a steady contribution below 70 km. The HF absorption is well correlated with solar X-ray flux, whereas VLF signals reflected below 90 km show very predictable diurnal and seasonal variations. The study of ionospheric absorption by George (1971) demonstrated that the HF absorption at low and middle latitudes may be geomagnet-



ically controlled, and that there is an "equatorial anomaly", i.e. a maximum at latitudes of about  $20^\circ$  north and south of the equator. This maximum, derived from vertical incidence absorption measurements, was attributed to enhanced D-region electron densities at these latitudes. Since the D-region anomaly coincides with the F-region equatorial anomaly, it has been suggested (Pradhan and Shirke, 1978) that the enhanced D-region ionization is caused by downward transport of nitric oxide produced in the F-region. Dyson and Bennet (1979) have recently pointed out that the deduced enhanced absorption may be too great owing to the neglect of a magnetoionic correction term. However, rocket measurements by Mechtly et al. (1967) seem to confirm the existence of the anomaly.

In conclusion, the quiescent low latitude D-region is becoming adequately described by the empirical formulae for HF absorption and VLF propagation that are used in current prediction schemes. From a physical point of view, the question of the size and cause of the equatorial anomaly in absorption needs a definite answer.

### 5.3 The Mid-Latitude D-Region

Because of its great variability, the mid-latitude D-region presents much more difficult problems for the forecaster than the ionosphere at lower latitudes. There are three main causes for this variability:

- (a) Energetic particle precipitation may, for several days following a magnetic storm, cause ion production exceeding that of the daytime solar X-ray and UV radiation.
- (b) The concentrations of minor constituents, important for the ionization balance such as NO, O, O<sub>3</sub>, O<sub>2</sub>(<sup>1</sup> $\Delta_g$ ) and H<sub>2</sub>O may be determined not only by purely aeronomic interactions, but also by transport mechanisms involving winds, wave motions and turbulence.
- (c) The ion composition, and hence the loss rate of ionization, depend strongly on the state (P, $\rho$ ,T) of the neutral atmosphere which may be highly perturbed in winter.

It is well established that there is a "storm aftereffect" causing enhanced mid-latitude D-region absorption for several days following major geomagnetic storms, and Larsen et al. (1977) have demonstrated that energetic electrons (10 - 200 keV) precipitate at subauroral latitudes during such conditions. This implies that modelling of D-region absorption effects at midlatitudes must either await adequate modelling of geomagnetic storms, or the development of a real time monitoring system to provide the details of energetic electron precipitation around the oval of impact.

Work in recent years (see for example Dynamical and Chemical coupling of the neutral and ionized atmosphere, eds. Grandal and Holtet 1977) has shown that the D-region ionization densities at middle latitudes are partly determined by the state of the neutral atmosphere, as discussed in Section 5.1. This state may be influenced by dynamical coupling with the lower atmosphere,



(Gregory & Manson, 1969). The ion composition and the ionization loss rates depend strongly upon atmospheric temperature (Chakrabarty, 1979), and the concentration of minor ionizable constituents may depend upon horizontal and vertical transport mechanisms. Recent observations of the winter anomaly in ionospheric absorption in Europe have demonstrated that this particular phenomenon can be explained by meteorological conditions in the mesosphere (Offermann, 1977; Thrane et al., 1979). The lower ionosphere also shows changes associated with stratospheric warmings. While the dynamical conditions within the stratosphere and mesosphere may differ significantly with altitude, latitude, and longitude during warmings, it is likely that the associated ionospheric changes will receive an explanation in terms of meteorological conditions also.

It is to be noted that the meteorological conditions referred to above are aspects of the dynamics of the winter circulation, and that these latter involve all altitudes from the surface to  $\approx 100$  km. Thus planetary and gravity waves generating in the troposphere can propagate to mesospheric and lower thermosphere altitudes; and may transport energy and momentum. The resulting circulation then determines, in fact, the transport of minor constituents. Again, interactions of standing and travelling planetary waves are believed to be the key to the "sudden warming" as it is observed in the Northern stratosphere.

We must conclude that our ability to predict conditions of ionization in the mid-latitude D-region, particularly during the winter months, November-March (and May-September in the Southern hemisphere) will depend upon our ability to predict pressures, densities, temperatures and winds in the mesosphere and lower thermosphere on useful latitude and longitude scales. At present, compilations of mean values of these variables, by latitude and month, are available, but longitudinal data are not available. Maps are routinely drawn to a pressure level of 0.4 mb; and less frequently to an altitude of 80 km. Statistics of perturbations of these variables have yet to be assembled. In particular, the determination of vertical motion will depend on development of new techniques. Finally, progress in respect to D-region studies, e.g. in modelling, is dependent on advances in the dynamical meteorology of the mesosphere and stratosphere, e.g. in respect to momentum budgets. Additional problems of the mid-latitude D-region, which may be linked to transport or temperature effects, await solution. For example, we do not understand the marked diurnal asymmetry observed in HF and MF absorption (Lastovicka 1978) and its seasonal variation.

#### 5.4 The High Latitude D-Region

Highly variable ion production caused by energetic particle precipitation is the major factor characterizing the high latitude D-region, though the meteorological effects discussed in the previous section are also likely to be significant. There are three major types of disturbances of importance in the high latitude D-region,

- (a) the excess ionization associated with precipitation of auroral electrons causing HF radio wave absorption and VLF phase perturbations in regions approximately along the auroral zones.
- (b) ionization produced over the entire polar cap above magnetic latitudes of  $60^\circ$  (PCA, polar cap absorption events) during Solar Proton Events (SPE).
- (c) relativistic electron precipitation events (REP) occurring at shell

values from  $L = 4-10$ .

There is extensive literature discussing these phenomena from a scientific point of view, (see Thrane 1978 and references therein), but standard prediction methods only take into account high latitude effects in a very rudimentary way. For example the "Bluedeck" prediction programme introduces a lump parameter "excess system loss at high latitudes", and the Omega predictions also include the average auroral and polar D-region effects on VLF phase velocity. It is not surprising that these schemes often fail in polar regions.

Auroral absorption occurs frequently in and near the auroral zone and is highly variable in time and space. In a statistical sense the phenomenon is well mapped using riometers (Relative ionospheric opacity meters) (see for example Hargreaves 1969). However, the application of these results to useful prediction schemes is not straightforward.

Recent work has been directed toward improving models of high latitude ionospheric parameters including the characteristics of one-way, vertical absorption at 30 MHz (Riometer absorption) (Vondrak et al). The model of auroral absorption includes the dependence on Kp value, but it is important to recognize that it was necessary to convert from the global Kp value to local values of K in order to adequately model this phenomenon. This latter point indicates that it is necessary to include and understand the short-term spatial variability of auroral absorption. Polar cap absorption events rarely occur near sunspot minimum, but may occur about once per month near solar maximum. The larger of these events cause pronounced D-region effects over the entire polar caps above  $60^\circ$  magnetic latitude and last for 1-12 days. The effects have been observed down to 30km altitude and are major sources of navigational errors when using Omega over arctic paths.

Attempts have been made to model the D-region effects of two large SPE's. Because of the relative spatial uniformity and slowly varying temporal features of SPE's as compared to auroral phenomena, successful modelling of the former is more likely to occur before the latter. Based on satellite spectral measurements of the energetic solar particle input to the ionosphere during the intense SPE of August 1972 and incoherent-scatter radar measurements of the resulting electron densities, Reagan and Watt (1976) derived effective recombination rates ( $\Psi_e$ ) as a function of solar zenith angle and ion-production rate. Swider et al (1978) have performed a detailed ion and neutral chemistry modelling of the 2-5 November 1969 SPE that is consistent with extensive rocket measurements of the in situ plasma and neutral atmosphere parameters, and with satellite measurements of the input ionization parameters.

If the energy, temporal and spatial variations of the precipitating particles are known from satellites, deposition codes are in existence (see for example Reagan and Watt, 1976) that provide the resulting ionization production profiles,  $q$ , to a sufficient degree of accuracy provided that appropriate consideration of the neutral atmosphere density variations with latitude and season are included. Based on the effective recombination rates obtained during SPE's and substantiated by chemistry modelling (Swider et al, 1978) electron density profiles can be obtained from the relationship given in Section 5.1.



The effective recombination rates obtained during daytime SPE's have been shown to be independent of the production rate over the range  $10^2$  to  $3 \times 10^5$  ion pairs/ $\text{CM}^2\text{-Sec}$  (Ulwick, 1972; Reagan and Watt, 1976). The effective recombination rates derived during moderate-to-severely disturbed conditions show less variation than during quiescent and mildly disturbed conditions (for example compare the data of Larsen et al 1977 and that of Reagan and Watt, 1976) and hence this simple relationship is probably more valid for the former cases.

During REP events which occur mainly in the morning hours, ( $L = 4,10$ ) electrons with energies of several hundred keV cause significant ionization down to altitudes of about 50 - 60 km. REP events (Bailey and Pomerantz (1965), Thorne (1974), Thorne and Larsen (1976)) are associated with geomagnetic storms, but no physical model adequate for prediction purposes exists. Because of the deep penetration of the particles, the REP events are important both for HF and VLF propagation. Observations (Doherty 1971) show the character of this type of event at LF.

Items which complicate the modelling of the high latitude disturbed D-region are:

- 1) the complex chemistry in the sunrise and twilight transition regions, 2) the effects of the long polar night on the ion and neutral chemistry, 3) the fact that intense events can change the composition of key minor neutral species such as ozone (Weeks, 1972, Heath et al 1978), nitric oxide and atomic oxygen, and 4) the large temporal, spatial and spectral changes that exist in energetic electron precipitation around the approximate auroral oval. We must conclude that current models are limited for useful D-region prediction at high latitudes.

## 5.5 Solar Flare Effects On the D-Region

X-rays emitted from solar flares cause increased D-region ionization over the entire sunlit hemisphere, affecting all three geographical regions discussed in the previous subsections. The sudden Ionospheric Disturbances (SID's) thus produced are observed as short wave fade outs (SWF) and Sudden Frequency Deviations (SFD) in the HF band, and as Sudden Phase Anomalies (SPA) in the VLF band. In the LF band the effects observed in both signal phase and amplitude suggest certain aspects that are related to both the VLF band and the MF and HF bands (Doherty, 1979). A great amount of work has been done to map and understand these effects which may be detrimental both for communications and navigation systems (see for example the series of papers by Deshpande and Mitra 1972)

Detailed knowledge of the X-ray spectrum outside the atmosphere and the atmospheric density profile are needed to give accurate ion production rate profiles in the lower D-region. From X-ray measurements below 10A, absolute X-ray spectra can be derived or predicted with sufficient accuracy (Deshpande et al, 1978). The ion production rates may, therefore, be calculated using the neutral density model atmosphere. The ionization loss rate has been observed to decrease during major flares. This decrease is probably due to both an increase in neutral temperature and an increase of X-ray ion production of  $\text{O}_2^+$  relative to the production of  $\text{NO}^+$  directly from the H-Lyman  $\alpha$ , as H Lyman  $\alpha$  fluxes do not increase appreciably during flares. (Chakrabarty and Chakra-



barty, 1979. Furthermore, the decrease in loss rate is greater the stronger the flare radiation. Thus, increased  $O_2^+$  production rates and decreased ionization loss rate considerably enhance the effects on radio wave propagation. The difficulty is that the ionization loss rate dependence upon X-ray flux is not available nor is the state of the D-region before the flare, which is important in controlling the response of the medium, except for unusually strong flares. Mitra and Deshpande (1972) have developed flare models of electron density profiles and HF absorption indicating a possible system for SID predictions. Ion composition measurements during flares and reaction rates of the ions and reactions involved (e.g. water cluster ions) are required for improvement in this area.

Solar X-ray flares are monitored on satellites and the information relayed to the forecasting center in real time. Experience shows that here certain empirical rules may be established relating X-ray gray body fluxes derived from satellite measurements in selected wavelength bands to (a) D-region propagation effects; and (b) ionization densities (Rothmuller, 1978; Deshpande, 1979). However in very few cases are both electron density and flare X-ray fluxes available for the same event. More such measurements are, therefore, needed. Such empirical rules may be the only practical approach for the near future. Combining the modelling for flare time and empirical relationships between X-ray fluxes and propagation parameters is possibly a first order system for prediction that can be developed with little effort. This also emphasizes the need of routine ground based SID monitoring by some standardized multi-frequency system covering SPA at VLF, phase and field strength at LF and absorption at MF, HF to improve the basis for prediction (Doherty, 1979; Deshpande and Mitra, 1972)

## 6. THE USER'S NEEDS

The atmospheric physicist may delight in mapping the intricate processes governing the formation of the ionospheric D-region, but the average user should not be expected to take an interest in them. He needs a timely, practical and, if possible, simple and inexpensive answer to his problems. We may assume that the users fall into two categories a) those dealing with long-term frequency planning and allocation, system coverage assessment or calibration through publication of phase delay tables, and b) communicators and navigators who face the day by day, hour by hour problems of adapting to the whims of the ionosphere. For those of the first group whose interest is in the low and mid-latitude ionosphere, present long term prediction schemes are useful but their deficiencies still limit performance. The needs here are for improved predictions during day-night transitions, at high latitudes and for propagation crossing the auroral zones and polar caps. Such schemes should meet the primary needs of the second group of users as for detection of events and time for alerting. Forecasts would be nice if reliable, but timely warnings may be essential.

It is essential to recognize that the ultimate user, the practicing navigator or short wave operator, can only use information which he receives in a timely fashion; he will only use information which has generally proved reliable in the past. Dissemination techniques on telemetry limit the timeliness of information. If it is necessary to provide published tables to navigators at sea, years may be required. Many forecasters do not seem to realize

priority TWX messages can take over  $24^h$  to reach their destinations. Since forecasts are often not reliable  $24^h$  in advance, users may disregard all forecast information. From the user's view, timely alerting to events which have already commenced is most desirable. If telemetry methods were available to support this, they would also provide a natural method to foster increased use of forecasts as these become more reliable.

The notion of an alert implies reliable detection and classification of events, i.e., categorizing a set of input variations as being due to, say a PCA, as opposed to an SID or equipment malfunction. At or shortly after the commencement of an event, forecasts for its subsequent effects, duration, etc., can often be quite accurate.

A general indicator of the level of solar activity, furnished for each day, seven days in advance would be useful and informative for the VLF/OMEGA user. Ideally, accurate predictions of the onset of M and X class flares would be disseminated to VLF/OMEGA users three to seven days in advance. However, even a rough projection of solar (X-ray) activity, expressed as a probability could be incorporated in the weekly OMEGA status report or special notification messages, given that the forecast accuracy is  $\lambda \leq 80\%$ . Statistical tests should be performed and results compiled on the accuracy of the three day projections of solar activity, and, in particular, proton events from the various forecast centers.

The user is also interested in information he can understand which relates to his needs. At some point in the overall prediction forecasting and alerting process the geophysical information needs to be applied to specific users. Navigators would like to know the magnitude of induced errors in miles, communicators want MUF's and LUF's. It is an essential requirement that prediction schemes be adapted to particular groups of users. For example, an aircrew operating in arctic regions and changing its position rapidly, will have different requirements from those of a shipborne operator. The operator could use not only indications of which frequency bands are most likely to serve him, but also instructions on alternative procedures such as relaying when these bands fail.

Another area which is rapidly becoming more important is that of HF data links. For such services predictions should estimate the effects of blackouts, fading, multipath and noise level upon the error rate for particular channels. Is it possible to estimate the magnitude of the errors produced by such effects, and to warn of their occurrence?

In general one should keep in mind that prediction requirements will depend upon the complexity and degree of sophistication of the communication or navigation system in use. For many users, for example in developing countries, simple schemes to be used on desk or even pocket calculators may suffice.

## 7. MAJOR SHORTCOMINGS OF PRESENT D-REGION PREDICTION TECHNIQUES AND POSSIBILITIES FOR FUTURE IMPROVEMENTS



In the preceding sections most of the important areas where present D-region predictions are inadequate have been mentioned, and they need only be briefly reviewed here.

At middle latitudes the major problem is coping with the great variability in D-region absorbing and reflecting properties due to particle precipitation and meteorological effects. This variability is particularly pronounced in winter. There is little hope of developing a complete physical model for prediction purposes in the foreseeable future. However, in certain areas improvements may be made. For example Spjeldvik and Lyons (1979) indicate possibilities for predicting ion production due to particle precipitation from the radiation belts during and after magnetic storms. They have used plasmaspheric ELF hiss as an indicator and base their arguments on casual relationships rather than statistical correlations. The paper offers hope for quantitative predictions of D-region after storm effects.

The winter anomaly in HF absorption is an accompaniment of the dynamical perturbations of the winter mesosphere circulation. It tends to occur in groups of days and has typically a geographical extent of a few thousand km. Careful groundbased continuous monitoring of the absorption may provide basis for predictions of the onset and duration of a period of winter anomaly absorption. Meteorological satellites are currently determining radiances in the mesosphere from which temperatures may be deduced, and also concentrations of minor constituents, e.g. Nimbus G. These observations offer possibilities for observing the conditions which cause the excess absorption.

The absorption in the upper D-region which is most important in HF communications varies approximately as the inverse square of the effective frequency  $1/(f+f_L)^2$  ( $f_L$  is the gyrofrequency about the longitudinal component of the earth's magnetic field) and is therefore relatively difficult to measure accurately, when using riometers for which  $f$  is normally around 20-30 MHz. The combination of modern solid state techniques with the interpretation methods developed by George (1971) could enable this absorption to be measured cheaply and easily using pulse absorption techniques at frequencies between about 2-5 MHz. The data base which exists at present is dangerously inadequate for practical applications and more data of this type will be required, in particular for the Middle Atmosphere Program. Thus it appears that both practical and scientific needs could be satisfied economically if such equipments could be produced and deployed in the near future.

The problem of predicting high latitude propagation conditions is very difficult, both on a long and short term basis. Some predictions in current use do not take proper account of, for example, the auroral absorption. On this point however, improvements have been made and are being implemented. In a CCIR Document, Supplement to Report 252-2, a new prediction scheme is described that includes a statistical distribution of auroral absorption in time and space. The report will be printed in the CCIR documents from the last General Assembly in Kyoto 1978. The computer programme has been developed at the Appleton Laboratory in the UK and is called Applab 3. Tests have shown that this programme is able to deal with high latitude absorption in a better way than previous schemes.



The auroral absorption is based on an analysis of riometer data by Foppiano (1975). This analysis includes sunspot number amongst the predictors, but not  $K_p$ , whose influence is much stronger.  $K_p$  is included in a more recent model by Vondrak et al (1977). The Institute for Telecommunication Sciences in Boulder has also developed an improved computer programme, IONCAP. These efforts will certainly improve high latitude long term predictions for high frequency waves. From the practical point of view it appears most likely that in the future satellite techniques for instantaneously mapping particle phenomena over the whole polar caps could provide the best techniques for forecasting radio propagation conditions in these zones. Considerable development is necessary before this becomes practical.

Predictions for VLF divide into long term predictions over periods of several years as used for coverage assessment and calibration, and warnings of events in progress. Short term forecasts are little used. Long term prediction for Omega propagation corrections is a major application of ionospheric D-region prediction. Current results are useful but introduce significant errors. Increased knowledge of the quiescent ionosphere including its spatial and temporal variations is needed. Greater understanding of and improved predictions for sunrise and sunset variations, sub auroral propagation, equatorial propagation and polar propagation are needed.

The forecast or warning of auroral substorm propagation effects is a particularly interesting problem where any progress in our understanding would be potentially very useful. Perhaps the important progress will come from mapping and monitoring of the sun and processes outside the earth's atmosphere, but there are certain possibilities for using ground-based techniques, such as magnetic and riometer recordings, and earth-observing photometric X-ray mapping techniques from satellites, that should not be neglected. Satellite imaging of the X-ray Bremsstrahlung that is emitted from the atmosphere in association with energetic electron precipitation offers a real potential of measuring simultaneously, the spatial extent and the variations of the hard precipitation around the entire oval. The technique which has been demonstrated by Imhof et al., (1974,1978) can be used to convert the measured X-ray spectra into the precipitating electron spectra within a resolution cell. Combining the spatial and spectral data with models of the effective recombination rates described in Sec. 5.4, the resulting enhanced electron densities around the oval and hence the magnitude of the HF absorption can be predicted. In the midnight and pre-midnight sectors the rapid poleward motion of the onset of a magnetic substorm seems to originate at some fairly definite latitude, about  $L=5.5$ , though with  $K_p$  variation (Hargreaves et al 1975, Jelly 1970). Thus it should be possible to get some warning of substorm commencement at a high latitude from observations at lower latitudes on the same meridian. The propagation of the onset in latitude and longitude has also been studied and average pictures presented, though there is a large and unexplained variation from case to case. (Hargreaves 1969, Hargreaves and Berry 1976.) Using such data it should be possible to get tens of minutes of warning for day-time activity, given observations in the night sector. Recent work (see Higbie et al Preprint No. 71 in these proceedings) has shown the  $\sim 30$  keV electron pitch angle anisotropies as observed at synchronous altitude may act as a sensitive indicator of the buildup of stresses in the outer magnetosphere. The development of such stresses is evidenced in the pre-midnight sector by the formation

of field-aligned anisotropies one to two hours prior to the onset of the expansion phase of a substorm. This indicator may prove to be an important one to two hour alert of a pending substorm. Monitoring of the interplanetary magnetic field (IMF) is one very effective indicator of a pending substorm. Caan et al (1977) examined 18 clear events for which the IMF turned southward after being directed northward for at least two hours, and found that such turnings were followed in about one hour by substorms as determined from magnetograms recorded at nightside auroral or midlatitude stations. There is a possibility (Hargreaves, 1979) that spikes in riometer absorption observed at the start of auroral absorption substorms may have prediction value, but further work on the statistics is needed. Weak absorption activity very often occurs some tens of minutes before the substorm begins in the night sector, and this weak activity shows a slow equatorward motion. Unfortunately it is usually weak and so any attempt to use this characteristic to predict the likely occurrence of a substorm could only be based on data of the highest quality. There is a strong possibility however, that auroral absorption substorms are nearly always preceded by other activity in the night sector and do not appear "out of the blue" from perfectly quiet conditions. There could be a basis for predictions here.

Ondoh and Obu (1979) have pointed out that HF signals propagating over polar paths show significant increases in signal strength as much as 10 hours before the onset of a substorm.

Polar cap absorption events caused by solar protons precipitating over the entire polar caps are important events from a practical point of view. The actual short time forecasts of the occurrence of PCA's must be based upon solar observations. Observations of a characteristic "U" shape in the 10 cm microwave radiation from the sun has proven to be a highly effective indicator of pending large solar particle events at the earth with an advance warning of 1-2 hours. (Smart and Shea, 1979). Even if we know the energy spectrum and spatial distribution of the proton fluxes, we cannot with certainty predict the D-region effects, though Bakshi and Barron (1979) present formulae predicting the absorption from characteristics of the solar radio burst. There are also some gross empirical relationships relating absorption (Reid 1972) and VLF phase (Westerlund et al 1969) to proton flux. There is a possibility as found by Japanese work (Kikuchi, private communication) that the magnitude and duration of SPE and geomagnetic storm effects on VLF (Omega signals) may be estimated after the onset of the disturbance from the statistical characteristic of propagation effects over several paths.

Prediction for solar flare condition holds promise. The basis for prediction can be further improved by in situ measurements of electron density covering more flare events, a few measurements on ion composition to provide a check for theoretical models, and by experimental and theoretical work oriented towards developing multifrequency systems of SID monitoring to give better quantitative information (Doherty, 1979; Deshpande et al., 1972).

Firmly established relations between solar phenomena and terrestrial weather could also form a basis for future D-region predictions. Obviously, there is a need for further research on the high latitude ionosphere. Until a better physical understanding has been achieved, predictions and forecasts may be improved through semiempirical relations.



## D-REGION PREDICTIONS REFERENCES

- Arnold, F. and Krankowsky, D. (1977): Ion composition and electron and ion loss processed in the earth's atmosphere. In: Dynamical and chemical coupling of neutral and ionized atmosphere. D Reidel Co, Dordrecht Holland. Eds: B. Grandal & J. A. Holtet pp 93-128
- Bailey, D. K. and Pomerantz, M. A. (1965): Relativistic electron precipitation into the mesosphere at subauroral latitudes. J. Geophys. Res., 70 pp 5823.
- Bain, W. C. (1974): The use of VLF propagation results in ionospheric modeling. In: ELF-VLF Radio Wave Propagation, D Reidel Publ. Co, Dordrecht-Holland, Ed: J. A. Holtet pp 151-163
- Bain, W. C. (1972): Model ionosphere for D-region at Summer noon during sunspot maximum. Proc IEE 119, pp 790-796
- Bain, W. C. (1974): Use of LF & VLF propagation data in studies of electron density profile in mid-latitudes. Indian Journ Radio & Space Phys 3 pp 113-118
- Bakshi, P. and Barron, W. R. (1979): Prediction of riometer absorption from solar flare radio burst characteristics. Volume IV, Solar-Terrestrial Predictions Proceedings (in press)
- Barghausen, A. F. et al (1969): Predicting long term operational parameters of high frequency skywave telecommunication systems. ESSA Tech Report ERL 110-ITS 78 US Gvt Printing Office
- Bobovnick, P. (1977): The effective recombination rate of the lower ionosphere determined from radio measurements at sunset. Contr Geophys Inst Sov Acad Sci 8, pp 59-61
- Bradley, P. A. (1975): Long turn JF propagation predictions for radio circuit planning. The Radio and Electronic Engineer 45, pp 31-41.
- Burgess, B. and Jones, T. B. (1975): The propagation of LF and VLF waves with reference to some system applications. The Radio and Electronic Engineer, 45, pp 47-61
- Caan, M. N. , McPherron, R. C. and Russell, C. T. (1977): Characteristics of the association between the interplanetary magnetic field and substorms. J. Geophys Res 82:4837
- Chakrabarty, D. K. (1977): Role of atomic oxygen and ozone in the D-region during disturbed conditions. In: Cospar Space Research XVIII. Pergamon Press Oxford and New York Ed: M. J. Rycroff and A. C. Stickland pp 253-259



- Chakrabarty, D. K., Chakrabarty, P. and Witt, G.: A theoretical attempt to explain some observed features of the D-region. J Geophys Res In Press; Preprint available
- Chakrabarty, D. K., Chakrabarty, P. and Witt, G.: Effect of variation in temperature and nitric oxide density on ion clustering in mesopause region during winter anomaly. J Atm Terr Phys In Press Preprint Available
- Chakrabarty, D. K., and Mitra, A. P.(1974): Theoretical models of D-region electron density profiles under different conditions. Indian Journ of Radio & Space Phys 3. pp 76-86
- Chakrabarty, D. K. and Chakrabarty, P. (1978): An attempt to identify the obscured paths of water cluster ions buildup in the D-region. J Atm Terr Phys 40. pp 437-442
- Chakrabarty, D. K. and Chakrabarty, P. (1977): Some studies of the daytime D-region during polar cap absorption. J atm Terr Phys 39, pp 57-67
- Chakrabarty, P, Chakrabarty, D. K. and Saha, A. K. (1977): Solar flux estimated from electron density and ion composition measurements in the lower thermosphere. J Geophys Res 82 pp 3299-3303
- Chakrabarty, P., Chakrabarty, D. K., and Bjorn, L (1977): Solar activity variation of nitric oxide in the E-region and its implications. Uppsala Ionospheric Observatory Report UIO-SR-77-02
- Chakrabarty, D. K., Chakrabarty, P. (1973): Some studies of D-region electron density profiles Indian Journ of Radio & Space Phys 2 pp 211-218
- Cole, K. D. (1977): Energy balance of the atmosphere under the influence of a disturbed sun and magnetospheric processes. In: Dynamical and chemical coupling of neutral and ionized atmosphere. D reidel Publ Co Dordrecht Holland. Eds: B. Grandal & J. A. Holtet pp 203-216
- Cook, F. E. and McCue, C. G. (1975): Solar-terrestrial relations and short term ionospheric forecasting. The Radio and Electronic Engineer 45, pp 11-30
- Davies, K (1978): In: AGARD CP 238
- Davis, R. M. and Berry, L. A. (1977): A revised model of the electron density in the lower ionosphere. Technical Report TR 111-77 Defense Communications Agency
- Deshpande, S. D. and Mitra, A. P. (1972): Ionospheric effects of solar flares-III. The quantitative relationship of flare X-rays to SID's. J Atmos Terr Phys 34 pp 243-253
- Deshpande, S. D. (1979): D-region prediction for solar flare conditions.

- Doherty, R. H. (1971): Observations suggesting particle precipitation at latitudes below 40°. Radio Science 6pp 639-646
- Doherty, R. H. (1979): Unpredicted variations in D-region response to solar X-ray events. Volume IV, Solar-Terrestrial Predictions Proceedings (in press)
- Dyson, P. L. and Bennett, J. A.: General formulae for absorption of radio waves in the ionosphere. J Atm Terr Phys In Print.
- Elkins, T. J.: A model of auroral substorm absorption. AFCRL-72-0413 Environmental research papers No 404. AFCRL Hanscom Air Force Base Mass 01730
- Foppiano, A. J. (1975): A new method for predicting the auroral absorption of HF sky waves. CCIR Interim Working Party Document 3
- George, P. L. (1971): The global morphology of the quantity  $fN_{\text{min}}$  in the D- and E- regions of the ionosphere. J Atm Terr Phys 33 pp 1893-1906
- George, P. L. and Bradley, P. A. (1974): A new method of predicting the ionospheric absorption of high frequency waves at oblique incidence. ITU Telecommunication Journal Geneva, May 1974 pp 3-7
- George, P. L. and Bradley, P. A. (1973): Relationship between HF absorption at vertical and oblique incidence. Proc IEEE 120pp 1355-13601
- Gregory, J. B., and Manson, A. H. (1969): Seasonal variations of electron densities below 100 km at mid-latitudes-II. Electron densities and atmospheric circulation. J Atm Terr Phys, 31 pp 703-729
- Hargreaves, J. K. (1969): Conjugate and closely spaced observations of auroral absorption - I Planet Space Sci 17, pp 1459-1484
- Hargreaves, J. K. (1974): Dynamics of auroral radio absorption in the midnight sector- The movements of absorption peaks in relation to the substorm onset. Planet Space Sci 22 pp 1427-1441
- Hargreaves, J. K. and Berry, M. G. (1976): The eastward movement of the structure of auroral radio absorption events in the morning sector. Ann Geophys 32 pp 401-406
- Hargreaves, J. K. (1969): Auroral absorption of HF radio waves in the ionosphere: A review of results from the first decade of riometry. Proc IEEE 57 pp 1348
- Hargreaves, J. K., Chivers and Axford, W. I. (1974): The development of the substorm in auroral radio absorption. J Atmos Terr Phys 23 p 905
- Haydon, G. W., Leftin, M. and Rosich, R. (1976): Predicting the performance of HF skywave telecommunication systems. Office of Telecommunications Report 76-102

- Imhof, W. L., Nakano, G. H., Johnson, R. G. and Reagan, J. B. (1974):  
 "Satellite Observations of Bremsstrahlung From Widespread Energetic  
 Electron Precipitation Events" Journ Geophys Res 79 p565
- Imhof, W. L., Nakano, G. H. and Reagan, J. B. (1978): "Satellite Observations  
 of Impulsive Bremsstrahlung X-Ray Events Associated With Substorms"  
Journ Geophys Res 83 p 4237
- Jelly, D. H. (1970): On the morphology of auroral absorption during substorms.  
Can Journ Phys 46, p 33
- Knapp, W. and Schwartz, K. (1975): "Aids for the Study of Electromagnetic  
 Blackout" DNA 3944H
- Knight, P.: MF propagation: a wave hop method for ionospheric field strength  
 predictions. BBC Engineering 100 pp 22-34 BBC London
- Labitzke, K and Petzoldt, K. (1979): Planetary waves in the strato-  
 and mesosphere during the Western European Winter Anomaly Campaign  
 1975/76 and their relation to absorption. J Atm Terr Phys (submitted)  
 Preprint available
- Larsen, T. R., Reagan, J. B., Imhof, W. L., Montbriand, L. E., and Belrose, J. S.  
 (1976): A coordinated study of energetic electron precipitation and  
 D-region electron densities over Ottawa during disturbed conditions,  
J Geophys Res 81 pp 2200-12
- Lastovicks, J. (1976): A planetary wave manifestation in the ionospheric  
 radio wave absorption in November 1972 Gerlands Beitr Geophysik  
 Leipzig 85, pp 276-282
- Lastovicka, J. and Triska, P. (1976): Winter anomaly observations at the  
 Panska Ves Observatory (Czechoslovakie 1971-73. Studia geophys et geod  
 20 pp 404-408
- Lastovicka, J. (1976): The dependence of the ionospheric absorption at w775  
 kHz on the intensity of ionizing radiation- Ionospheric implications  
 Pure and applied geophysics 114 pp321-131
- Lastovicka, J. (1977): Seasonal variation in the asymmetry of diurnal  
 variation of absorption in the lower ionosphere. J Atm Terr Phys 39,  
 pp 891-894
- Lastovicka, J. (1975): SID-monitoring at the Panska Ves Observatory  
Geofysikalni Sbornik XXIII pp 315-330
- Lastovicka, J.: The effects of sector boundary crossings of the interplanetary  
 magnetic field at 0-100 km altitudes in winter Preprint available
- Lauter et al: HHI-STP reports 9 and 10



- Lee, M. K. and Nisbeth, J. S. (1975): Propagation predictions and studies using a ray tracing program combine with a theoretical ionospheric mode. IEEE Trans Ant and Prop p132
- Lucas, D. L. and Haydon, G. W. (1966): Predicting statistical performance indexes for high frequency telecommunication systems. ESSA Tech Report IER 1-ITSA 1 US Printing Office
- Lukkari, L and Kangas, J. (1977): Correlation electron precipitation and magnetic IPDP events near the plasmapause J Geophys Res 82 pp 4750-4756
- Manson, A. H., Gregory, J. B., Meek, C. E., and Stephenson, D. G. (1978): Winds and wave motions to 110 km at mid-latitudes V. An analysis of data from September 1974- April 1975 J Atm Sci 35 pp 594-599
- McNamara, L. F. (1978): Ionospheric D-region profile data base. World Data Center A Report UAG-67
- Mechtly, E. A., Rao, Skaperdas, and Smith, L. G. (1969): Latitudinal variation of the lower ionosphere. Radio Science 4 pp 517-520
- Morris, P. B., and Cha, M. Y. (1974): Omega propagation corrections: Background and computational algorithm. Report ONSOD-01-74 Dept of Transportation, US Coast Guard, Omega Navigation System Operations Detail
- Naval Ocean Systems Center, San Diego CA 92152 (1978): The NOCS real-time ionospheric radio propagation assessment capability. TD 119
- Offermann, D. (1977): Some results from the European Winter Anomaly Campaign 1975/76 Dynamical and Chemical Coupling of Neutral and Ionized Atmosphere Ed: B Grandal & J A Holtet. D Reidel Publ Co Dordrecht Holland pp 235-252
- Oyinloye, J. O. (1979): Prediction of radio wave absorption in the ionosphere. Volume IV, Solar-Terrestrial Predictions Proceedings (in press)
- Ondoh, T. and Obu, K. (1979): Prediction of HF communication disturbances by pre-SC HF field increases on polar paths crossing the auroral zone. Volume IV, Solar-Terrestrial Predictions Proceedings (in press)
- Pradhan, S. N. and Shirke, J. S. (1978): Geomagnetic control of mesospheric nitric oxide concentrations from simultaneous D and F region ionization measurements. Ann Geophys 34 pp 125-131
- Ranta, H. and Ranta, A. (1977): Daily variation of absorption in the D-region using riometer data at high latitudes. J Atm Terr Phys 39, pp 309-312
- Ranta, H. and Ranta, A. (1977): Study of latitudinal, diurnal and seasonal variation of ionospheric absorption according to observations of the riometer network in Finland. Geophysica 14, 2, pp 182-198

- Ranta, Hilka: The onset of an auroral absorption substorm. J Geophys Res  
In Press Preprint available
- Ranta, H. and Ranta, A.: Riometer measurements of ionospheric radio wave  
absorption. J Atm Terr Phys In Press Preprint available
- Rawer, K. (1977): Mid and high latitude reference ionosphere. In: Dynamical  
and Chemical Coupling of Neutral and Ionized Atmosphere Eds: B. Grandal  
and J. A. Holtet D Reidel Publ Co Dordrecht Holland pp 129-144
- Reagan, J. B. (1977): Ionization processes In: Dynamical and Chemical  
Coupling of neutral and Ionized Atmosphere Eds: B. Grandal and J. A.  
Holtet. D Reidel Publ Co Dordrecht Holland pp 145-160
- Reid, G. C. (1972): A review of ionospheric radio propagation effects  
associated with solar proton events. Proceedings of COSPAR Symposium  
on Solar Particle Event of November 1969. Ed: James C. Ulwick  
AFCRL-72-0474 pp 201
- Rothmuller, I. J. (1978): Real time propagation assessment AGARD CP 238
- Samuel, J. C. and Bradley, P. A. (1975): A new form of representation of  
the diurnal and solar cycle variations of ionospheric absorption.  
J Atm Terr Phys 37 pp 131-141
- Schultz, L. D. and Gallet, R. M. (1970): A survey and analysis of normal  
absorption measurements. ESSA Professional Paper 4.
- Swanson, E. R. and Brown, R. P. (1972): Omega propagation prediction primer,  
Technical Note 2101, -Naval Electronics Laboratory Center
- Swanson, E. R. (1977): Propagation effects on Omega, AGARD Conf Proc on  
Propagation limitations of navigation and positioning systems. AGARD  
CCP-209
- Swanson, E. R. and Levine, P. H. : Automated Omega/VLF monitoring and  
forecasting for air traffic safety enhancement. A progress report.  
Preprint available.
- Swanson, E. R., Larsen, T. R. and Thrane, E. V. : Factors affecting Omega  
accuracy NDRE Rept 71 1978 Norwegian Defence Res Est N-2007 Kjeller,  
Norway
- Swanson, E. R. (1972): VLF phase prediction, VLF-propagation: Proc VLF-  
symposium, Sandefjord, Norway, 27-30 October 1971, Report 7201, Norwegian  
Institute of Cosmic Physics 8.1 to 8.36
- Swider, W. and Dean, W. A. (1975): "Effective Electron Loss Coefficient  
of the Disturbed Daytime D Region" J Geophysical Research 80 pp 1815

- Swider, W., Keneshea, T. J. and Foley, C. I. (1978): An SPE-disturbed D-region model. Planet Space Sci 26 pp 883.
- Thorne, R. M. (1974): A possible cause for dayside relativistic electron precipitation events. J Atmos Terr Phys 36 pp 635
- Thrane, E. V. : Geophysical disturbances and their predictability. AGARD Conference Proc 1978 In Press
- Thrane, E. V., Bangert, W., Beran, D., Friedrich, M., Grandal, B. , Hagen, O., Loidl, A., Spenner, K., Schwentek, H., Torkar, K. M., Ugletveit, F. : Ion production and effective electron loss rate in the mesosphere and lower thermosphere during the Western Europe Winter anomaly campaign 1975-76 Submitted to J Atm Terr Phys
- Vondrak, R.R., Smith, G., Hatfield, V. E., Tsunoda, R. T. , Frank, V. R. and Perreault, P. D. (1977): Chalanika model of the high-latitude ionosphere for application to HF propagation prediction. SRI Report or contract F19628-77-c-0102
- Vondrak et al (1977): "Chatanika Model of the High Latitude Ionosphere for Application to HF Propagation Predictions" RADC-TR-78-7
- Westerlund, S., Reder, F. H., Abom, C. : Effects of polar capaabsorption events on VLF transmissions Planet Space Sci 17 p 1329
- Zevakina, R. A., Lavrova, Y. V. Liakhova, L. N., (1967): Principles for predicting ionospheric and magnetic disturbances and short range radio services Nauka, Moscow

#### Members of the Working Group D-Region Predictions

Dr. F. Arnold  
Max Planck Inst. für Kernphysik  
69 Heidelberg, Postfach 1249  
WEST GERMANY

Dr. W. C. Bain  
Appleton Laboratory  
Ditton Park  
Slough SL3 9JX, ENGLAND

Prof. Dr. G. Becker  
Physik. Tech. Bundesanstalt  
Abteilung 1  
Bundesallee 100  
3300 Braunschweig, GERMANY

Dr. H. A. von Biel  
Physics Dept  
University of Canterbury  
Private Bag  
Christchurch, NEW ZEALAND

Mr. L. Bjørn  
Uppsala Ionosfareobservatorium  
S-755-90 Uppsala, SWEDEN

Dr. Asgeir Brekke  
Nordlysobservatoriet  
9001 TROMSØ, NORWAY



Dr. Boyd Burgess  
Radio and Navigation Department  
Royal Aircraft Establishment  
Farnborough, Hants GU14 6TD  
ENGLAND

Dr. D. K. Chakrabarty, (Room 773)  
Physical Research Laboratory  
Navangpura  
Ahmedabad 380009, INDIA

Dr. A. D. Danilov  
Institute of Applied Geophysics  
107258 Glebovskaya Str 20 b  
Moscow, USSR

Dr. S. D. Deshpande  
Department of Applied Physics  
College of Engineering  
AMRAVATI 444603, INDIA

Dr. R. H. Doherty  
Colorado Research and Prediction Lab.  
P. O. Box 1065  
Boulder, Colorado 80302, USA

Dr. A. J. Ferraro  
The Pennsylvania State University  
318 Electrical Engineering East  
University Park  
Pennsylvania 16802, USA

Dr. G. J. Fraser  
Physics Dept.  
University of Canterbury  
Private Bag  
Christchurch, NEW ZEALAND

Dr. M. Friedrich  
Inst. für Nachrichtentechnik  
und Wellenausbreitung  
Technische Universität Graz  
Inffeldgasse 12, A-8010 Graz, AUSTRIA

Dr. J. B. Gregory  
Physics Department  
University of Saskatchewan  
Saskatoon STN OWO, CANADA

Dr. J. K. Hargreaves  
Environmental Science Dept.  
University of Lancaster  
Bailrigg, Lancaster LA1 41Q  
UNITED KINGDOM

Dr. T. B. Jones  
Department of Physics  
University of Leicester  
Leicester LE1 7RH, ENGLAND

Dr. I. H. Keroub  
Radio Observatory  
p. o. Box 4655  
Haifa, ISRAEL

Dr. Takashi Kikuchi  
Inubo Radio Wave Observatory  
Radio Research Laboratory  
Tennodai 9961, Choshi  
Chiba, JAPAN 288

Professor K. Labitzke  
Freie Universität Berlin  
Institut für Meteorologie  
Podbielskiallee 62  
1 Berlin 33, GERMANY

Dr. T. R. Larsen  
Norwegian Defence Res. Establ.  
p.o. Box 25  
N-2007 Kjeller, NORWAY

Dr. H. S. Lee  
The Pennsylvania State University  
318 Electrical Engineering East  
University Park  
Pennsylvania 16802, USA

Dr. Paul Levine  
Chief Scientist  
Megatek Corporation  
1055 Shafter St.  
San Diego, California 92106, USA

Dr. C. G. McCue  
Ionospheric Prediction Service  
Department of Science  
P.O. Box 702  
Darlinghurst NSW 2010, AUSTRALIA

Dr. L. F. McNamara  
Ionospheric Prediction Service  
Department of Science  
P.O. Box 702  
Darlinghurst NSW 2010, AUSTRALIA

Dr. Peter Morris  
Omega Navigation Systems Operation  
Detail (ONSOD)  
Headquarters US Coast Guard  
(G-ONSOD/43)  
Washington, DC 20590, USA

Professor M. Nicolet  
External Geophysics  
Brussels University  
30, Avenue Den Doorn  
1180 Brussels, BELGIUM

Professor D. Offermann  
Gesamthochschule Wuppertal  
Gausstrasse 20  
56 Wuppertal1, WEST GERMANY

Dr. J. O. Oyinloye  
Department of Physics  
University of Ibadan  
Ibadan, NIGERIA

Dr. W. R. Piggott  
British Antarctic Survey  
Radingley Road  
Cambridge CB3 0ET, ENGLAND

Mrs. Hilikka Ranta  
Geophysical Observatory  
SF-99600 Sodankyla, FINLAND

Dr. J. B. Reagan  
Lockheed Palo Alto Research Lab.  
3251 Hanover Street  
Palo Alto, California 94304, USA

Dr. Kristian Schlegel  
Max Planck Inst. für Aeronomie  
Postfach 20  
D-3411 Katlenburg - Lindau 3  
WEST GERMANY

Dr. Warren A. Schlueter  
Mission Research Corporation  
735 State Street  
Post Office Drawer 719  
Santa Barbara  
California 93102, USA

Dr. Peter Stauning  
Det Danske Meteorologiske Institut  
Ionosfaerelaboratoriet  
Danmarks Tekniske Højskole  
2800 Lyngby, DENMARK

Dr. E. Swanson  
Naval Electronics Laboratory Center  
271 Catalina Boulevard  
SanDiego  
California 92152, USA

Dr. W. Swider  
Air Force Geophysics Laboratory  
Hanscom Air Force Base  
Massachusetts 01731, USA

Dr. Kurt Taenzer  
Standard Elektrik Lorenz AG  
Dept CNS/TWB  
Postfach 400749  
D-7000 Stuttgart 40, GERMANY

Dr. E. V. Thrane (Chairman)  
Norwegian Defence Research Estab.  
Division for Electronics  
P.O. Box 25 - N-2007 Kjeller, Norway

## ION PRODUCTION IN THE D-REGION

William Swider  
Composition Branch (LKD)  
Air Force Geophysics Laboratory  
Hanscom AFB, MA. 01731

Ionization processes for the D-region are reviewed. Levels of ionization range from the fairly steady rates produced through galactic cosmic rays to the large rates generated during solar flares and associated geomagnetic storms. The normal quiescent daytime D-region is ionized by solar UV radiation available because of certain "windows" and solar X rays, although this latter source is generally weak. Cosmic rays are important roughly below 65 km. At night, stellar x-ray sources may be important along with precipitating electrons at high latitudes from the quiet magnetosphere. Solar flares can result in ionization rates due to X rays, electrons or nuclei which are greatly enhanced over quiet conditions. Typical ionization production rates from these various sources are illustrated.

### 1. Introduction

There are a number of different sources of D-region ionization. These sources can be highly variable and quite inhomogeneous with respect to both space and time. Galactic cosmic rays are the steadiest component, having no diurnal variation but some latitudinal variation. However, this source is basically a minor one for the D-region except below about 65 km, where it becomes the prominent quiet source. The principal ionization agent for the quiescent daytime D-region is still thought to be HLy- $\alpha$  radiation which ionizes nitric oxide and has little variation with solar activity but, of course, has an intensity which is dependent upon the local solar zenith angle since it is absorbed by molecular oxygen with a unity optical depth near 75 km for an overhead sun. A typical peak ion-pair production rate as a result of  $\text{HLY}\alpha + \text{NO} \rightarrow \text{NO}^+ + e$  is of the order of  $1 \text{ cm}^{-3} \text{ s}^{-1}$ , depending upon the NO distribution itself, which appears to be subject to important changes with latitude and/or season. Of somewhat less prominence, the ionization of  $\text{O}_2$  ( $^1 \Delta$ ) as a consequence of minor "windows" for EUV radiation, particularly one at  $1108\text{\AA}$ , produces  $\text{O}_2^+$  ions and electrons in the D-region. Solar X rays at 1-10  $\text{\AA}$  provide a minor contribution to the ionization of the D-region during low solar activity. However, at high solar activity they are comparable to  $\text{HLY}\alpha + \text{NO} \rightarrow \text{NO}^+ + e$  as a source of ion-pairs. During solar flares, X rays at these wavelengths may become the dominant D-region ionization agent creating an enhanced D-region over a period of roughly an hour. Like  $\text{HLY}\alpha + \text{NO}$ , their local influence depends upon the local solar zenith angle since x-ray absorption is by means of the principal atmospheric constituents,  $\text{O}_2$ ,  $\text{N}_2$  and A. At night, distant stellar objects are thought to be



weak ionization mechanisms for low latitudes.

At high latitudes, precipitating particles may cause substantial increases in the D-region ionization production rates. Auroras may penetrate into the upper D-region and solar proton events (polar cap absorption events) enhance the entire D-region within the polar caps. These latter events possess ionization production rates which are as much as four orders of magnitude in excess of the quiet daytime rates. Among the other ionization agents in the D-region are a narrow electron drizzle band on the equatorward side of the auroral ovals and relativistic electron precipitation events. Bremsstrahlung radiation, generated in conjunction with precipitating electrons, may be important because even though this emission is generally weak, it penetrates into the low D-region where there are few ionization mechanisms.

## 2. Ionization by Galactic Cosmic Rays

Ionization of the D-region as a result of cosmic rays was introduced by Nicolet (1958). Nicolet and Aikin (1960) observed that this ionization source is important in the lower D-region, below about 65 km. There is evidence (e.g., Bain et al., 1972) that a persistent minor layer exists near this altitude which should be more prominent at low solar activity rather than high solar activity and which should be more pronounced at high rather than low latitudes, although precipitating particles at high latitudes may obscure this latter fact. The layer, not always evident, has been dubbed the C-layer.

The behavior of the ionization created by galactic cosmic rays is very different from that generated by solar ionization sources. The galactic particles, mainly protons, are shielded somewhat more from the Earth during high solar activity since the solar interplanetary magnetic field is greater. This influence is not too strong, however, leading only to about a factor of two difference in the ionization created during low, as opposed to high, solar activity. More significant is the latitudinal control exerted on the galactic cosmic rays by the terrestrial magnetic field. Thus, there is roughly a factor of ten reduction in ionization starting from polar latitudes ( $\lambda_m \approx 60^\circ$ ) to the magnetic equator  $\lambda_m = 0$ . Below the "knee" ( $\lambda_m \approx 60^\circ$ ) in the cosmic ray penetration ability, the cutoff rigidity for cosmic rays increases like  $\cos^4 \lambda_m$ . However, this is only a rough picture since the "knee" has been observed to disappear in one year, 1954, of low solar activity (Neher and Anderson, 1962). Further, solar events can lead to temporary (Forbush) decreases in the incoming galactic rays.

Empirical formulas for the ionization of the D-region by galactic cosmic rays can be developed, as <sup>was</sup> most recently done by Heaps (1978) but the simple tabulation of Swider (1969) may still suffice for D-region ionization purposes since at the altitudes of interest, the ionization rate due to cosmic rays is proportioned to the atmospheric density with an ionization coefficient  $J_0$  ( $s^{-1}$ ).

Ionization coefficients may be extrapolated from the following table roughly in terms of a sunspot number of 200 for high solar activity and a number 0 for low solar activity.

Table 1. GALACTIC IONIZATION COEFFICIENT  $10^{18} \times J_O(s^{-1})$  AS A FUNCTION OF HIGH AND LOW SOLAR ACTIVITY FOR THE INDICATED GEOMAGNETIC LATITUDES (SWIDER, 1969)

$\lambda_m^O$	0	10	20	30	40	50	60	70	80	90
$J_O(\text{low})$	3	3	4	6	9	14	19	21	22	22
$J_O(\text{high})$	2	2	2	3	5	8	9	9	9	9

Energetic particles, be they X rays, electrons, protons or heavier nuclei, create one ion-pair for approximately every 35 eV absorbed in air. Most of the ionization is actually produced by secondary electrons generated in the absorption process. The ionization is essentially distributed in proportion to the constituents present. Thus, Dalgarno (1967), noting a 4:1 ratio for  $N_2:O_2$  and adapting the same ratio for dissociative ionization, obtained 0.64:0.16:0.16:0.04 for the relative ion formation rate  $q(N_2^+):q(N^+):q(O_2^+):q(O^+)$  with  $q(e)$  being their sum. A more detailed estimate, since there are some differences in the cross sections for the various processes as a function of the constituent ( $N_2$  or  $O_2$ ), has been given by Swider (1969), 0.62:0.17:0.14:0.07. He has further concluded that since ionic reactions occur rapidly at D-region altitudes, these relative ionic production rates may be simplified to  $q(O_2^+):q(NO^+)$  with a ratio of 9:1.

### 3. UV IONIZATION OF MINOR CONSTITUENTS

#### 3.1 NO Ionization

By chance, the strong solar hydrogen Lyman-alpha emission line at 1216 $\text{\AA}$  coincides with a UV absorption "window" in the atmosphere. This radiation is absorbed only by molecular oxygen with a cross section of  $10^{-20} \text{ cm}^2$ . Unity optical depth,  $[O_2]H 10^{-20} = 1$ , occurs at approximately 75 km for an overhead sun, solar zenith angle  $\chi = 0$ , where  $O_2$  is the molecular oxygen concentration in  $\text{cm}^{-3}$  and  $H$  is the scale height of this constituent in cm. The scale height at D-region altitudes is basically the same for all major gas species since the atmosphere is well mixed at these levels. The ionization coefficient ( $s^{-1}$ ) is a maximum at unity optical depth.

Nicolet (1945) was first to point out that  $H\text{Ly}\alpha$  radiation (10.2eV) could ionize any NO present in the D-region since this constituent has an ionization potential of 9.27 eV in contrast to  $O_2$  and  $N_2$  which have ionization potentials of 12.06 and 15.58 eV, respectively. Although it is one of the oldest ideas extant concerning the source of the D-region,  $H\text{Ly}\alpha + NO \rightarrow NO^+ + e$  is still currently believed to be the major ionization mechanism of the quiescent daytime D-region,

$$q(NO^+) = 6 \times 10^{-7} [NO] \exp \left\{ -[O_2] H 10^{-20} Ch(X, \chi) \right\} \quad (1)$$

where  $Ch(X, \chi)$  is the Chapman function which equates to  $\sec \chi$  for  $\chi \lesssim 75^\circ$  with D-region scale heights (Swider, 1964), the parameter  $X$  being a function of  $H$  and distance from the Earth's center. The ionization coefficient,  $6 \times 10^{-7} s^{-1}$ , is the product of the ionization cross section for NO at 1216 $\text{\AA}$ ,  $2 \times 10^{-18} \text{ cm}^2$ , and the nominal line intensity,  $3 \times 10^{11} \text{ photons/cm}^2\text{-s}$ . There is little change in the intensity with solar activity. Swider (1969)

suggested  $(3-5) \times 10^{11} \text{ cm}^{-2} \text{ s}^{-1}$  and more recently Nicolet (1978) gave  $(2-4) \times 10^{11} \text{ cm}^{-2} \text{ s}^{-1}$  for the variation in the photon intensity of this line with solar activity.

The largest uncertainty in the ionization production rate computed from equation (1) arises from our uncertainty in the nitric oxide concentration,  $[\text{NO}]$ . This gas apparently increases with latitude and solar activity (Swider, 1978) in the lower thermosphere. The relationship with solar activity is shown in Figure 1 as based upon  $[\text{NO}]$  profiles derived by Swider (1978) from ion composition data excluding high latitude winter results. Golshan and Sechrist (1975) have reported a similar behavior for peak  $[\text{NO}]$  derived in their study of E-region ion composition data. Less information is known concerning  $[\text{NO}]$  at key D-region altitudes 65-85 km. Swider (1972) has demonstrated that a D-region mixing ratio of  $10^{-8}$  for  $[\text{NO}]/[\text{M}]$ , where  $[\text{M}]$  is the total atmospheric gas concentration, will yield an electron distribution equal or greater to five typical  $[e]$  profiles measured by Mechtly and Smith (1968). Certainly, a better knowledge of  $[\text{NO}]$  is critical to any model of the quiet daytime D-region world-wide.

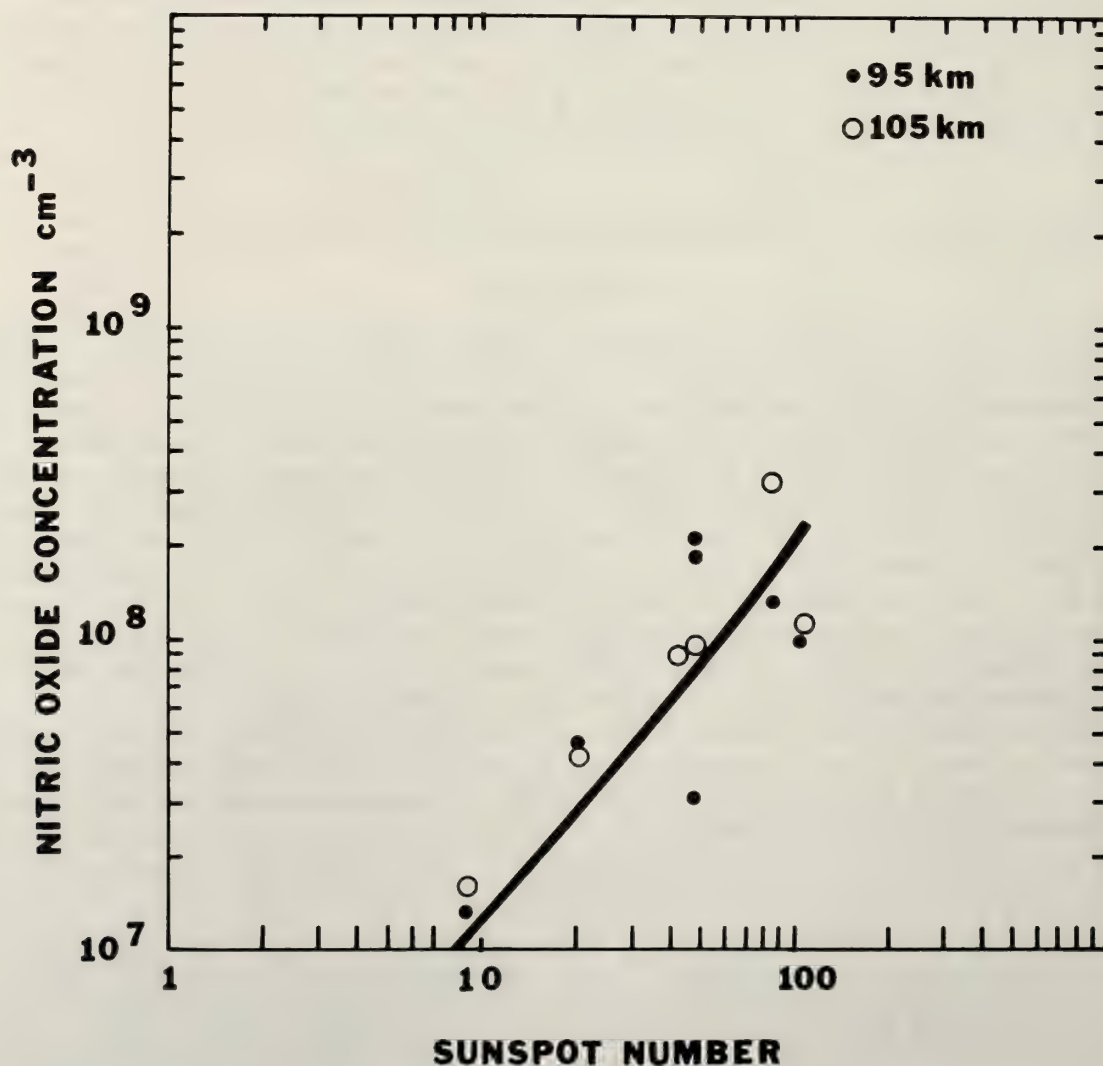


Figure 1. THE VARIATION OF NITRIC OXIDE WITH SOLAR ACTIVITY



Enhancements of nitric oxide at high latitudes in winter, perhaps as a result of transport effects, may be a, if not the, cause of the phenomenon called "winter anomaly." Certainly, there is evidence that NO is indeed enhanced in the winter D-region (e.g., Swider, 1978). This gas is apparently enhanced at auroral latitudes as a result of auroral electron precipitation and the subsequent chemical interactions (Rusch and Barth, 1975; Swider and Narcisi, 1977).

Solar HLy $\alpha$  radiation undergoes multiple scattering as a result of the hydrogen in the terrestrial atmosphere. Models for this scattering effect have been developed out to 12 Earth radii (Meier and Mange, 1973). The scattered radiation is insignificant in the daytime as compared to the direct radiation but may make some contribution at night. Strobel et al. (1974) have modelled the scattered HLy $\alpha$  photon intensity in terms of solar zenith angles from 90 to 180°. However, they stress that their model is not accurate below 110 km, being based upon an extrapolation of observations obtained with rocket-borne instruments. At 80 km for  $\chi = 90^\circ$ , they give an effective scattered HLy $\alpha$  intensity of  $3 \times 10^9 \text{ cm}^{-2} \text{ s}^{-1}$ . This value is about 1 per cent of the daytime direct flux value at  $\chi = 30^\circ$ . The scattered flux decreases with increasing  $\chi$ .

### 3.2 O<sub>2</sub> (<sup>1</sup>Δg) Ionization

The discovery of significant concentrations of O<sub>2</sub> (<sup>1</sup>Δg) at D-region altitudes by Evans et al. (1968) prompted Hunten and McElroy (1968) to postulate O<sub>2</sub><sup>+</sup> formation by the ionization of this metastable constituent by UV continuum radiation at 1027-1118Å, the longer wavelength being the threshold value for O<sub>2</sub> (<sup>1</sup>Δg) ionization, the shorter wavelength being the threshold value for the ionization of O<sub>2</sub> in its ground state. In particular, about 50 per cent of the O<sub>2</sub> (<sup>1</sup>Δg) ionization results from a Si III multiplet which lies at the deepest window in this region at 1108Å (Paulsen et al. (1972). However, Hunten and McElroy (1968) omitted CO<sub>2</sub> absorption in this region which is important even though  $[\text{CO}_2]/[\text{M}] \sim 3 \times 10^{-4}$ , where  $[\text{M}]$  is the total gas particle concentration. The most accurate expression for the ionization of O<sub>2</sub> (<sup>1</sup>Δg) is that given by Paulsen et al. (1972) who averaged over 75 wavelength intervals and 9 solar lines. Their formula is

$$q(\text{O}_2^+) = [\text{O}_2(\text{}^1\Delta\text{g})] \left\{ 0.549 \times 10^{-9} \exp(-2.406 \times 10^{-20} [\text{O}_2] \text{HCh}(X, \chi)) + 2.614 \times 10^{-9} \exp(-8.508 \times 10^{-20} [\text{O}_2] \text{HCh}(X, \chi)) \right\} \quad (2)$$

As in the case for  $q(\text{NO}^+)$ , formula (1) (ionization of NO by HLy $\alpha$ ), the knowledge of the concentration of the ionizable constituent, in this case O<sub>2</sub> (<sup>1</sup>Δg), is crucial to the calculation. The distribution of this gas component, which has a one-hour radiative lifetime, is not well known yet as a function of season, latitude and/or solar activity. Thus, in general, while  $q(\text{O}_2^+) < q(\text{NO}^+)$ , there may be occasions when O<sub>2</sub> (<sup>1</sup>Δg) is enhanced thereby resulting in  $q(\text{O}_2^+) > q(\text{NO}^+)$ . It should be realized that above about 80 km, O<sub>2</sub> (<sup>1</sup>Δg) is formed principally from  $2[\text{O}] + \text{M} \rightarrow \text{O}_2(\text{}^1\Delta\text{g}) + \text{M}$  and (Llewellyn and Solheim)  $\text{OH} + \text{O} \rightarrow \text{O}_2(\text{}^1\Delta\text{g}) + \text{H}$ . The distribution of atomic oxygen is dependent upon both transport and photochemistry. Enhancements of O<sub>2</sub> (<sup>1</sup>Δg)

might be another factor contributing to the "winter anomaly" condition. There is information that both the  $O_2$  ( $^1\Delta_g$ ) and the NO concentrations were increased over normal conditions during recent winter anomaly observations over Europe (Offermann, 1978).

#### 4. X-RAY IONIZATION

##### 4.1 Non-flare Solar X rays

A number of workers (Nicolet and Aikin, 1960, Culhane et al., 1964; Poppoff et al., 1964) have summarized the behavior of solar X rays in the 1-10 Å wavelength interval which create ionization in the D-region. Swider (1969), for example, has subdivided the non-flare solar x-ray spectrum into three groups with total energy fluxes below 8 Å of  $10^{-3}$ ,  $9 \times 10^{-5}$  and  $8.2 \times 10^{-6}$  erg cm $^{-2}$  s $^{-1}$ . Satellite observations (Kreplin et al., 1962) reveal that a sudden ionospheric disturbance (SID) occurs whenever the energy flux below 8 Å exceeds  $2 \times 10^{-3}$  erg cm $^{-2}$  s $^{-1}$ . Ionization production rates for the  $\leq 8$  Å x-ray energy flux levels mentioned above are numerically tabulated by Swider (1969), who also lists all the appropriate x-ray cross sections, for various solar zenith angles and selected D-region altitudes. With increasing solar activity, X rays become comparable to HLy $\alpha$  + NO as an ionization source.

Ionization rates as a result of X rays are compared in Figure 2 to  $q(NO^+)$  and  $q(O_2^+)$  for a solar zenith angle of 45°. Ion-pair production rates resulting from galactic cosmic rays at 50° geomagnetic latitude are depicted also. Values for  $q(O_2^+)$  are taken from Paulsen et al. (1972). Ionization rates  $q(NO^+)$ , equation (1), were computed on the basis of the nitric oxide concentrations measured by Baker et al. (1977) which, incidentally, result in  $[NO]/[M]$  mixing ratios of about  $10^{-8}$ . Although their observations of the NO gamma bands yield nitric oxide concentrations which are somewhat on the low side of most such measurements, these  $[NO]$  values are not at all unreasonable when compared to ionospheric evidence (Swider, 1972; 1978).

The results depicted in Figure 2 illustrate that solar X rays for very quiet and quiet conditions may be ignored. It is only with barely quiet conditions that X rays are seen to be comparable with other ionization sources, and even then we must remember that the  $[NO]$  profile adopted for  $q(NO^+)$  is smaller than most workers have assumed in recent years for D-region studies. This conclusion is compatible with a study by Triska and Lastovicka (1970) involving a comparison of SOLRAD 9 x-ray data and radio absorption data by the A3 method recorded at Panská Ves Observatory. They estimated that there was no 1-8 Å x-ray control upon D-region absorption if the total energy was less than  $10^{-3}$  erg cm $^{-2}$  s $^{-1}$ . Some influence was evident if the energy flux was in the range  $1-5 \times 10^{-3}$  erg cm $^{-2}$  s $^{-1}$  whereas for a value exceeding about  $10^{-2}$  erg cm $^{-2}$  s $^{-1}$ , the radio absorption was controlled by X rays. These effects, reasonably enough, were apparently independent of the solar zenith angle except perhaps for very large  $\chi$ . As in the case for all high energy ( $> 100$  eV) photons or particles which lose energy in the



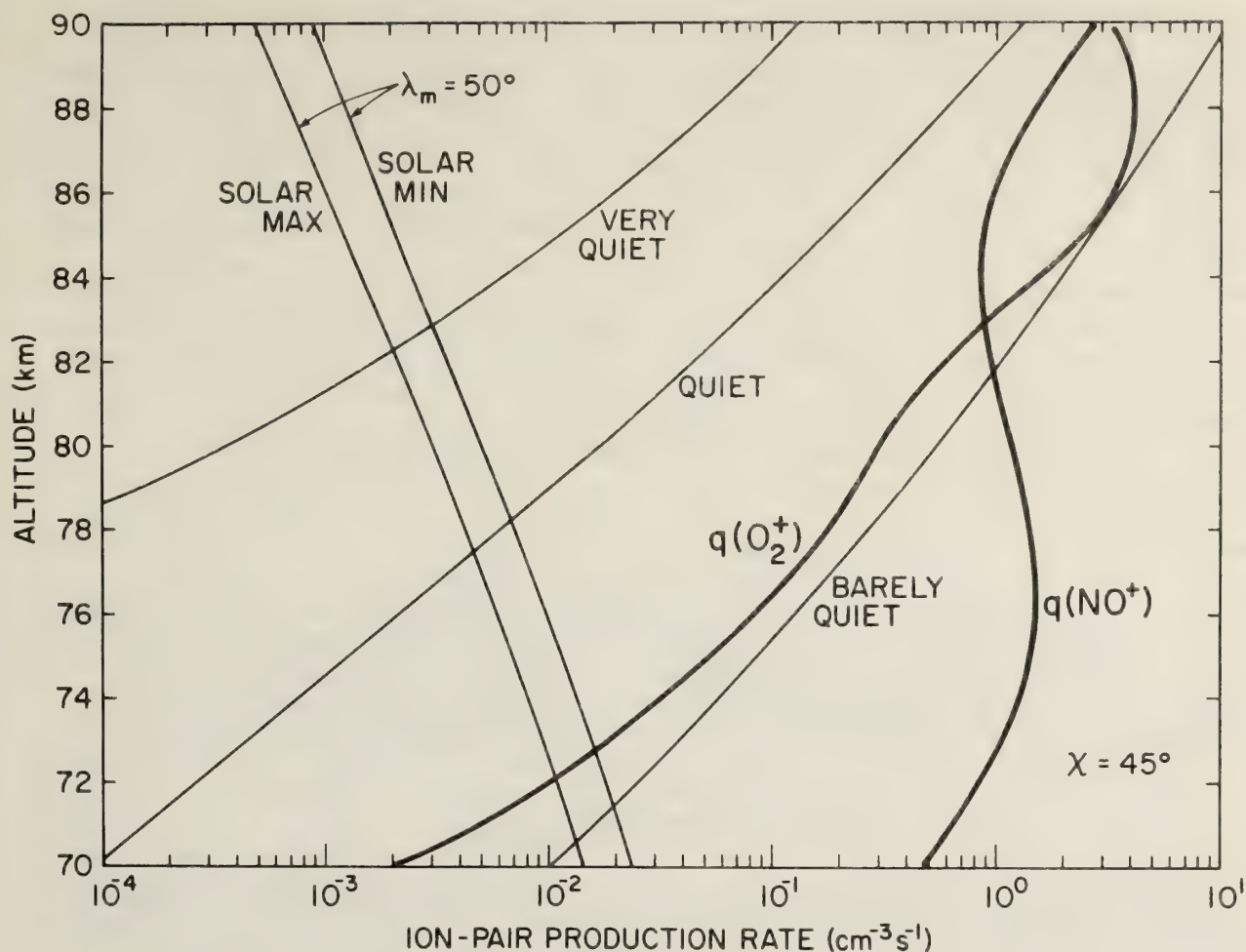


Figure 2. NON-FLARE X RAY IONIZATION RATES COMPARED TO OTHER IONIZATION RATES

D-region, the total column number of ion-pairs formed  $\text{cm}^{-2} \text{s}^{-1}$  is the total incoming energy  $\text{cm}^{-2} \text{s}^{-1}$  divided by 35 eV with the ionic production rate effectively distributed as 90%  $\text{O}_2^+$  ions and 10%  $\text{NO}^+$  ions as we have noted already in the section on galactic rays.

#### 4.2 Solar X Ray Flares

The brightening of the Sun during its various eruptions can lead to substantial increases in the solar x-ray output so that X rays become the dominant D-region ionization source. The mean duration of an x-ray flare event is about a half-hour. The effect on the ionosphere is to therefore generate an SID on this time-scale in the sunlit hemisphere.

The burst component of the solar x-ray flux at 2-12Å was analyzed by Drake (1971) for the period July 1966 to September 1968. The data were obtained with the Explorer 33 and 35 satellites. Some 4028 burst peaks were identified in this time frame, about half of which were suitable



for various statistical studies. Drake (1971) found that the most common rise time was 4 m, whereas the most common decay time was 10-13 m with the most common duration for an event being 16 m. Furthermore, he has provided various formulas for differential rise time, differential flux time, integral values, etc. The common times should be distinguished from the mean rise, decay and total duration times which appear to be more like 8, 20, and 30 minutes, respectively. These "soft" x-ray (2-12Å) portions of solar bursts/flares are related to the "slow" components of the overall solar event (Donnelly, 1976). Numerous other studies of solar x-ray flare spectrums exist. We only mention the work of Deshpande and Mitra (1972) who considered 9 events observed on OGO-4 and SOLRAD-9 because 5 other related papers were written concerning these flares and their effects on the D-region by one or the other, or both, of these authors. Most of these papers appear in the same journal issue as the work cited.

Since the cross sections for x-ray absorption are well-known (e.g. Swider, 1969), ionization models pertaining to the soft component of solar bursts may be easily determined if the temporal behavior of the X rays and their intensity are known from satellite observations or other means. Shimabukuro (1970), for example, has ascertained good temporal correlation between 3.3 mm and soft x-ray bursts. Solar flares will impact only upon the sunlit hemisphere of the earth, furthermore, their influence will wane the greater the solar zenith angle. Finally, it should be reemphasized that solar flares contribute to D-region ionization when the total energy flux exceeds about  $2 \times 10^{-3} \text{ erg cm}^{-2} \text{ s}^{-1}$  for  $\lambda \leq 8\text{Å}$  (Kreplin et al., 1962).

#### 4.3 Stellar X Ray Sources

Over the last decade, numerous authors have discussed whether or not certain celestial objects including Sco X-1, the strongest x-ray source in the Southern Hemisphere, Tau XR-1, Cen XR-2 and Cen XR-4 contribute to D-region ionization. Svennesson et al. (1972) determined that there were some stellar source effects upon long nighttime VLF paths. Although there appears to be much conflicting evidence, Baird and Francey (1972) suggest that the acceptance of the dominance of electron precipitation over celestial X rays for  $L \geq 2$  would remove most of these conflicts. They have noticed that all of the positive reports by Svennesson et al. (1972) are in conjunction with VLF paths with mid points at  $L \leq 1.7$ . We might further add in support of this premise that galactic cosmic ray ionization becomes more important with increasing latitude and [NO] increases with latitude making its nighttime ionization by scattered HLy more significant with higher latitudes.

Karszenbaum and Gagliardini (1975) illustrate that while X rays from Tau XR-1 are not a significant ionization source, those from Sco X-1 exceed the galactic cosmic ray source for  $\lambda_m = 0^\circ$  but not for  $\lambda_m = 55^\circ$  where  $\lambda_m$  is the terrestrial geomagnetic latitude. However, we caution that their ionization rates all appear to be on the high side (cf. Chilton and Crary, 1971). If Sco X-1 is subject to flares, the contribution of this ionization source to the D-region can be significant even for daytime conditions. However, Sastri and Murthy (1975) found the evidence to be inconclusive in investigating the effect of one Sco X-1 flare event upon the daytime D-region.

A not uninteresting facet of these studies is the conclusion of Svensson et al. (1972) that  $\text{NO} \leq 10^7 \text{ cm}^{-3}$  at 80 km since a higher value like  $3 \times 10^7 \text{ cm}^{-3}$  precludes a role for Sco XR-1 as a D-region ionization source. This low concentration is in harmony with the experimental [NO] profile of Baker et al. (1977) which was obtained by means of NO gamma band measurements and which was used in computing  $q(\text{NO}^+)$  in Figure 2, and it is in harmony with the [NO] profiles deduced by Swider (1978) from ionic composition data.

## 5. IONIZATION BY ENERGETIC PARTICLES

### 5.1 Ionization generated by solar protons and alpha particles

Energetic nuclei are ejected from the sun during solar flares characterized by type IV radio bursts having a U-shaped spectrum (Castelli et al., 1967; O'Brien, 1970; McCracken and Rao, 1970). These particles are mainly protons, but alpha particles are equally abundant sometimes (Biswas and Fichtel, 1965). Especially in the early phases of an event energetic electrons are likely to be important also (Hakura, 1965). If the earth lies in the path of these energetic particles, and if these particles are sufficient in number, a polar cap absorption event (PCA) occurs. (The terms solar proton event (SPE) and solar cosmic ray event (SCR) are also used.) There were 48 principal PCAs during the period 1952 to 1963 (Bailey, 1964; compare McDonald, 1963), based on 2.5-dB 30-MHz absorption at Fort Churchill, Canada. Zmuda and Potemra (1972) list 30 events of 1.0 dB absorption or greater for the four years 1965-1969. The number of events for this period is reduced to 14 if a 2.5 dB criterion is adopted. Smart and Shea (1971) have devised a classification system for solar proton events. They relate 30-MHz riometer absorption measurements to satellite observations of the quantity of protons with kinetic energies above 10 MeV.

Solar cosmic rays generally have much lower kinetic energies than their galactic ray counterparts. They deposit the bulk of their energy in the D-region and/or upper stratosphere. Figure 3 illustrates the relationship between the penetration depth of vertically incident protons, alpha particles (helium nuclei) and electrons as a function of their initial kinetic energy. The distribution of protons or alpha particles is most commonly fitted by an exponential rigidity spectrum or a power law spectrum in energy. Ionization rates have been computed by numerous researchers (e.g. Bailey, 1959; Reid, 1961; Freier and Webber, 1962; Adams and Masley, 1966). Most recently, Swider (1977) has illustrated electron-ion pair production rates generated by protons and alpha particles for both types of distributions with selected typical values of rigidity or appropriate powers of energy. Here, we illustrate (Figure 4) only the ion-pair production rates determined for the moderate SPE of 2-5 November 1969 and the intense SPE of 2-11 August 1972. These ionization rates are far in excess of those for quiet conditions (Figure 2) and represent the largest possible D-region ionization rates exclusive of nuclear bursts. Velinov (1970) has derived general expressions for SPE conditions. Duback and Barker (1971) claim to have simplified his formulas, but this is disputed by Velinov (1974).



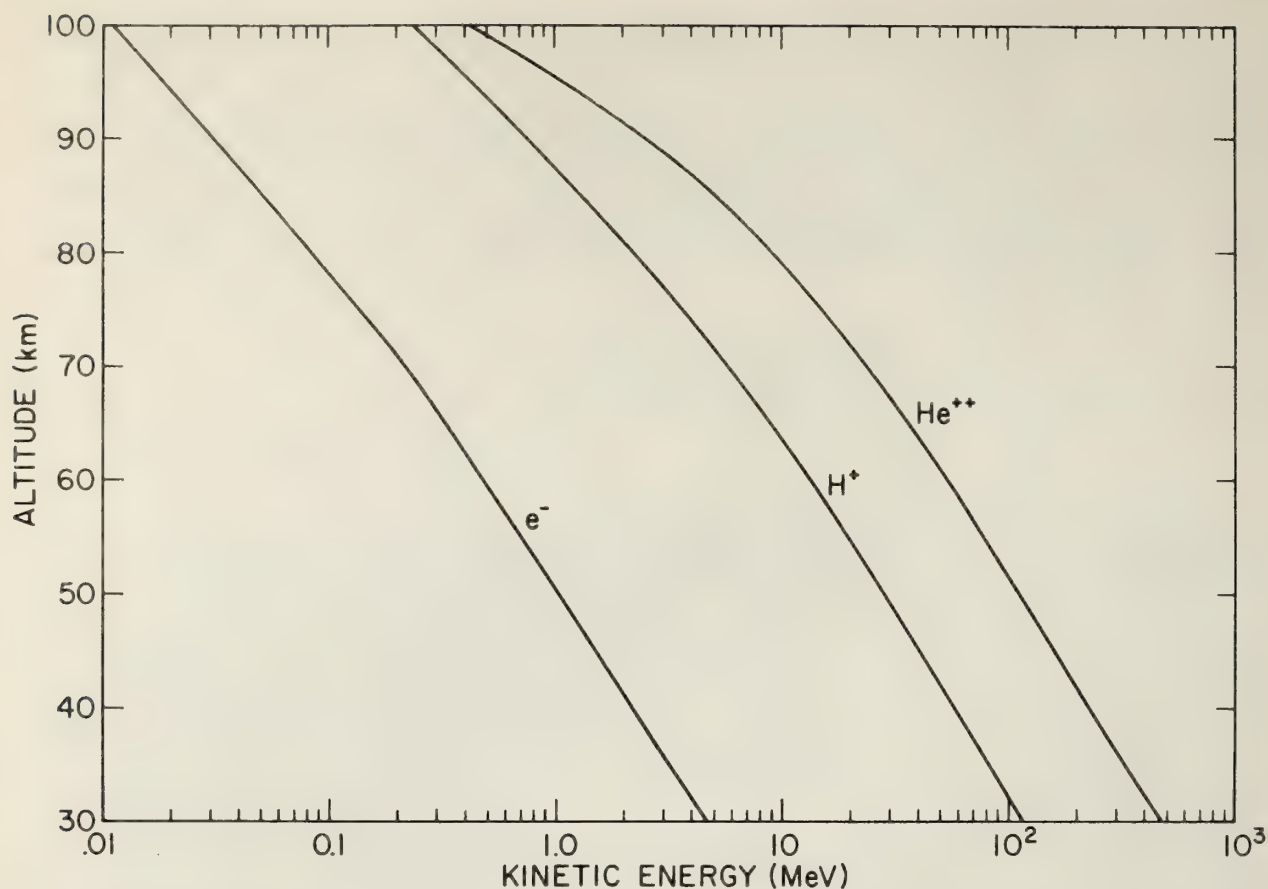


Figure 3. STOPPING ALTITUDE FOR VERTICALLY INCIDENT PARTICLES AS A FUNCTION OF ENERGY

Since solar cosmic rays are less energetic or "rigid" than galactic cosmic rays, the influence of the terrestrial magnetic field is more pervasive. Thus, while solar cosmic rays essentially enter the polar regions unhindered, they are rapidly cut off with decreasing geomagnetic latitude below this region. In general, the low energy cutoff tends to decrease at a given geomagnetic latitude during an SPE for latitudes as low as  $55^\circ$  corrected geomagnetic latitude with  $65^\circ$  being more normal (Hakura, 1965). As the intensity of the precipitating particles diminishes, the cutoff latitude recedes towards its pole with a slightly stronger recovery by day as compared to night (Leinback, 1967; Paulikas et al., 1968; Imhof et al., 1971). The typical SPE has a duration of several days.

## 5.2 Ionization Generated by Energetic Electrons

The contribution of energetic electrons to ionization production in the D-region is one of the most difficult subjects to assess since the temporal and spatial behavior of these energetic particles is so erratic and thus so difficult to elucidate. As we have already noted, energetic electrons may be important in the first few hours of an SPE (Hakura, 1965).



Auroral precipitation is generally confined to E-region altitudes. Calculations for the ionization created by the absorption of energetic electrons into the auroral E-region have been performed by a number of researchers (e.g. Rees, 1969; Berger et al., 1970; Strickland et al., 1976). D-region ionization is achieved only if the aurora involves electrons with kinetic energy in excess of about 20 keV (Figure 3) thus giving rise to an auroral absorption (AA) event.

Electron precipitation into the D-region also occurs which is not connected with visible auroras or SPE's. Whalen et al. (1971) have delineated a zone just equatorward of the auroral oval which contains a persistent, if weak, "drizzle" of energetic electrons into the D-region. Gough and Collin (1973) have concluded that at South Uist ( $L = 3.5$ ), energetic electrons are the dominant nighttime ionization source in the D-region  $15 \pm 10$  per cent of the time during solar minimum and  $35 \pm 2$  per cent of the time during solar

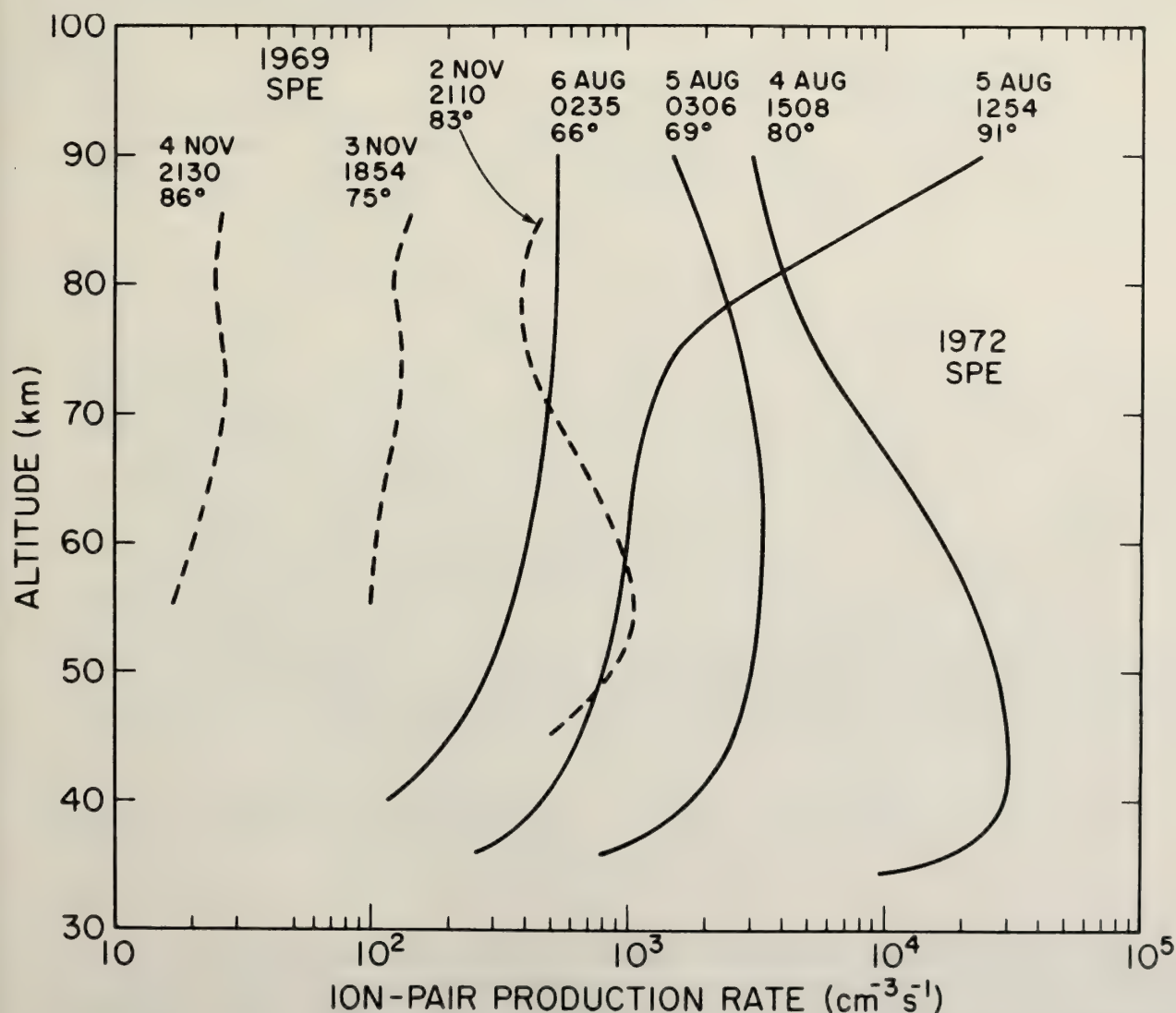


Figure 4. IONIZATION RATES FOR THE NOVEMBER 1969 SPE (ULWICK, 1973) AND THE AUGUST 1972 SPE (REAGAN AND WATT, 1976) FOR UT AND  $\chi$  INDICATED.

maximum. Significant enhancements in the ionization rates of the mid-latitude ( $L \approx 2-4$ ) D-region may occur during and especially following magnetic storms as discussed by Larsen et al. (1976) in conjunction with partial reflection data at Ottawa, Canada ( $L = 3.6$ ) and satellite data concerning precipitating electrons. Wratt (1976) has found that such electrons associated with the storm effects of the August 1972 SPE resulted in unusually high daytime  $[e]$  in the lower D-region at Christchurch, New Zealand ( $L = 2.6$ ). These geomagnetic storm related enlargements of the D-region often attain a maximum effect 2-4 days after storm onset and may last up to ten days according to work cited by Wratt (1976), the effects being observed in VLF, LF and MF radio wave phase measurements and in absorption measurements for MF and HF waves. The overall picture concerning particle precipitation at low and middle latitudes have been reviewed by Paulikas (1975). In his discussion of the South Atlantic anomaly he concluded that the aeronomic effects of particle precipitation were not well established and somewhat controversial. However, it appears that definite changes are observed in the D-region during or after large magnetic storms.

Other sporadic ionization associated with electrons is that resulting from REP (relativistic electron precipitation) events. They may occur as often as 10 per cent of the time (Thorne, 1977). Generally, these events are restricted to sub-auroral latitudes,  $4 < L < 8$ . Much work remains to be done on the statistics of their occurrences.

### 5.3 Ionization Generated by Bremsstrahlung

Loss of energy by the bremsstrahlung process is significant for particles when they reach relativistic energies. Thus, this x-ray ionization source must be considered during REP events (Thorne, 1977). However, even for auroral electrons, some bremsstrahlung will be created and it may be a significant D-region ionization source at times only because such radiation penetrates to low D-region altitudes where other ionization sources are weak. Kamiyama (1967), for example, computed maximum impact ionization rates for a  $2\pi \times 10^7 \text{ cm}^{-2} \text{ sec}^{-1}$  monoenergetic electron flux of  $1.7 \times 10^3 \text{ cm}^{-3} \text{ s}^{-1}$  at 116 km ( $E = 10 \text{ keV}$ ) and  $2.3 \times 10^4 \text{ cm}^{-3} \text{ s}^{-1}$  near 88 km ( $E = 100 \text{ keV}$ ) with concurrent maximum bremsstrahlung ionization rates of  $3.6 \times 10^{-2} \text{ cm}^{-3} \text{ s}^{-1}$  near 97 km ( $E = 10 \text{ keV}$ ) and  $1.9 \times 10^1 \text{ cm}^{-3} \text{ s}^{-1}$  near 45 km ( $E = 100 \text{ keV}$ ). An example of a bremsstrahlung ionization production rate distribution is shown in Figure 5 with other ionization results taken from a paper by Reagan (1977).

## 6. SUMMARY

We have seen that there is a number of D-region ionization sources, several of which are quite variable. Ionization by galactic cosmic rays in the D-region is fairly simple to determine but care must be taken with respect to the latitude of interest. The UV ionization of NO and O<sub>2</sub> ( $^1\Delta_g$ ) is easy to compute but these constituents themselves are subject to changes which are not well documented. X ray absorption cross sections are well

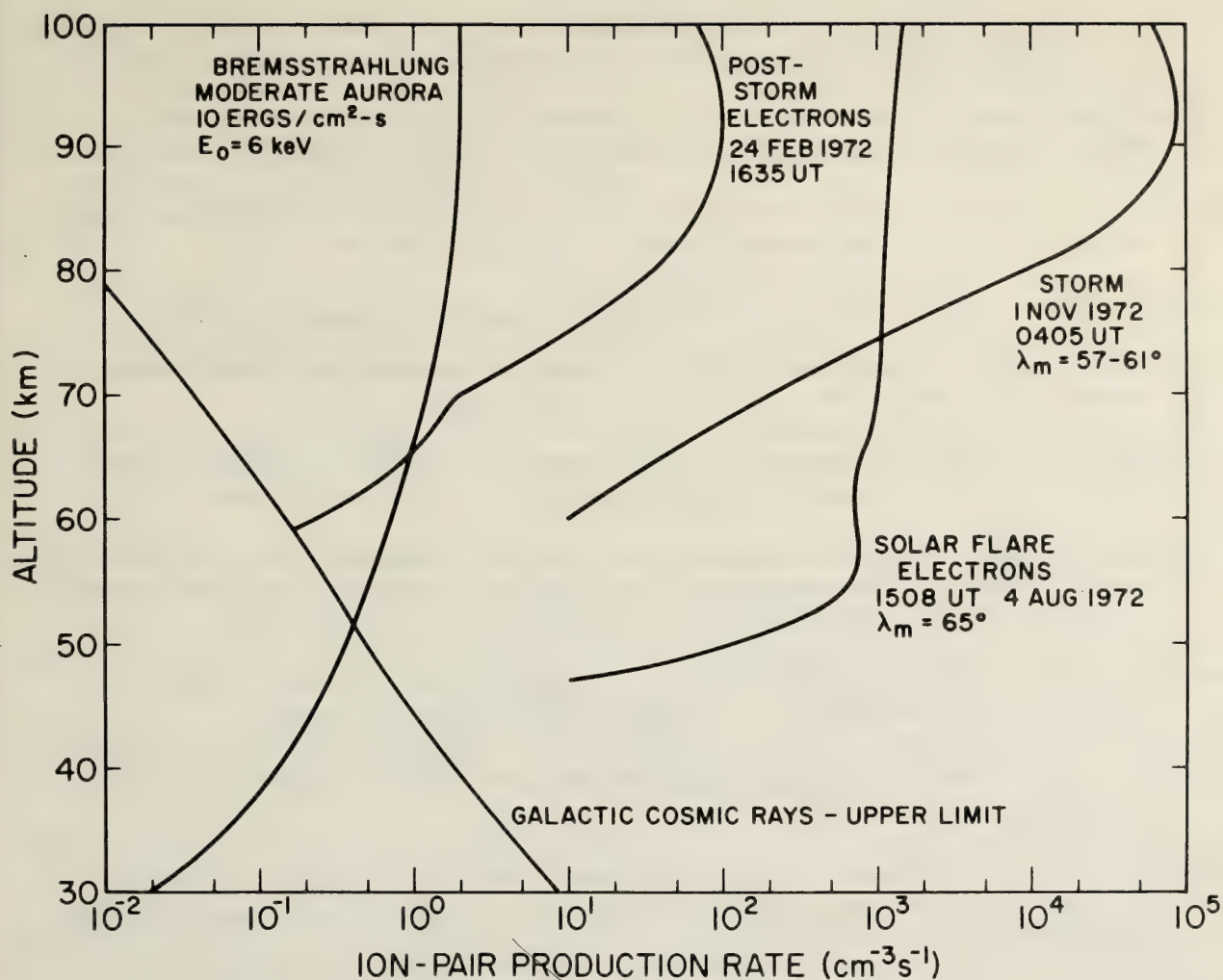


Figure 5. SELECTED ELECTRON-INDUCED IONIZATION RATES (REAGAN, 1977) COMPARED WITH COSMIC RAY RATES

established but solar x-ray intensities can vary appreciably. Similarly, while electron and proton absorption cross sections for air are known, the ionization associated with such particles when ejected from the sun or when precipitated by perturbations of the terrestrial magnetic field require a real-time flux measurement system like SELDADS (Williams, 1976) if the impact of energetic particles upon the D-region are to be properly assessed. The effect of nighttime sources like celestial X rays and scattered H $\gamma$  $\alpha$  radiation can be roughly estimated.



## REFERENCES

- Adams, G.W., and A.J. Masley (1966): Theoretical study of cosmic noise absorption due to solar cosmic radiation. Planet Space Sci., 14: 277.
- Bailey, D.K. (1959): Abnormal ionization in the lower ionosphere associated with cosmic-ray flux enhancements. Proc. IRE, 47:255.
- Bailey, D.K. (1964): Polar-cap absorption. Planet. Space Sci., 12:495.
- Bain, W.C., and M.D. Harrison (1974): Model ionosphere for D-region at summer noon during sunspot maximum. Proc. IEE, 119:790.
- Baird, G.A., and R.J. Francey (1972): Comments on the ionospheric detection of cosmic-ray phenomena. J. Geophys. Res., 77:1966.
- Baker, K.D., A.F. Nagy, R.O. Olsen, E.S. Oran, J. Randhawa, D.F. Strobel, and T. Tohmatsu. Measurement of the nitric oxide altitude distribution in the mid-latitude mesosphere. J. Geophys. Res., 82:3281.
- Berger, M.J., and S.M. Seltzer (1970): Energy deposition by auroral electrons in the atmosphere. J. Atmos. Terr. Phys., 32:1015.
- Biswas, S., and C.E. Fichtel (1965) Composition of solar cosmic rays. Space Sci. Rev., 4:709.
- Castelli, J.P., J. Aarons, and G.A. Michael (1967) Flux density measurements of radio bursts of proton-producing flares and nonproton flares. J. Geophys. Res., 72:5491.
- Chilton, C.J., and J.H. Crary (1971): VLF observations of nighttime D-region ionization enhancement by the Scorpius XR-1 x-ray source. Radio Sci., 6:699.
- Culhane, J.L., A.P. Willmore, K.A. Pounds, and P.W. Sanford (1964): Variability of the solar x-ray spectrum below 15Å. Space Res., 4:741.
- Dalgarno, A. (1967): Atmospheric reactions with energetic particles. Space Res., 7:849.
- Deshpande, S.D., and A.P. Mitra (1972): Ionospheric effects of solar flares - II. The flare spectrum below 10Å deduced from satellite observations. J. Atmos. Terr. Phys., 34:229.
- Donnelly, R.F. (1976): Empirical models of solar flare X ray and EUV emission for use in studying their E and F region effects. J. Geophys. Res., 81:4745.
- Drake, J.F. (1971): Characteristics of soft solar x-ray bursts. Solar Phys., 16:152.

- Dubach, J., and W.A. Barker (1971): Charged particle induced ionization rates in planetary atmospheres. J. Atmos. Terr. Phys., 33:1287.
- Evans, W.F.J., D.M. Hunten, E.J. Llewellyn, and A. Vallance Jones (1968): Altitude profile of the infrared atmospheric system of oxygen in the dayglow. J. Geophys. Res. 73, 2885.
- Freier, P.S., and W.R. Weber (1963): Exponential rigidity spectrums for solar-flare cosmic rays. J. Geophys. Res., 68:1605.
- Golshan, N., and C.F. Sechrist, Jr. (1975): Seasonal and solar cycle variation of E region nitric oxide. Radio Sci. 10:305.
- Gough, M.P., and H.L. Collin (1973): Energetic electron precipitation as a source of ionization in the night-time D-region over the midlatitude rocket range, South Uist. J. Atmos. Terr. Phys., 35:835.
- Hakura, Y. (1965): Initial phase of the PCA events and compositions of the solar cosmic radiations. J. Rad. Res. Lab. (Japan), 12:231.
- Heaps, M.G. (1978): Parametrization of the cosmic ray ion-pair production rate above 18 km. Planet. Space Sci. 26:513.
- Hunten, D.M., and M.B. McElroy (1968): Metastable O<sub>2</sub> (<sup>1</sup>Δg) as a major source of ions in the D-region, J. Geophys. Res. 73:2421.
- Imhof, W.L., J.B. Reagan, and E.E. Gaines (1971): Solar particle cutoffs as observed at low altitudes. J. Geophys. Res., 76:4276.
- Kamiyama, H. (1967): The electron density distribution in the lower ionosphere produced through impact ionization by precipitating electrons and through photoionization by the associated bremsstrahlung X rays. J. Geomag. Geoelect., 19:27.
- Karszenbaum, H., and D.A. Gagliardini (1975): Galactic x-ray sources and the ionospheric D-region. Nature, 257:34.
- Kreplin, R.W., T.A. Chubb, and H. Friedman (1962): X ray and Lyman-alpha emission from the sun as measured from the NRL SR-1 satellite. J. Geophys. Res., 67:2231.
- Larsen, T.R., J.B. Reagan, W.L. Imhof, L.E. Montbriand, and J.S. Belrose (1976): A coordinated study of energetic electron precipitation and D-region electron concentrations over Ottawa during disturbed conditions. J. Geophys. Res., 81:2200.
- Leinbach, H. (1967) Midday recoveries of polar cap absorption. J. Geophys. Res., 72:5473.

- Llewellyn, E.J., and B.H. Solheim (1978): The excitation of the infrared atmospheric oxygen bands in the nightglow. Planet.Space Sci., 26:533.
- McCracken, K.G., and U.R. Rao (1970): Solar cosmic ray phenomena. Space Sci. Rev. 11:155.
- McDonald, F.B. (Editor) (1963): Solar proton manual, NASA Tech. Rept. R-169, Off. Tech. Serv., Dept. Commerce, Washington, D.C. 20230.
- Mechtly, E.A., and L.G. Smith (1968): Seasonal variation of the lower ionosphere at Wallops Island during the IQSY. J. Atmos. Terr. Phys. 30:1555.
- Meier, R.R., and P. Mange (1973): Spatial and temporal variation of the Lyman-alpha airglow and related atomic hydrogen distributions. Planet. Space Sci. 21:309.
- Neher, H.V., and H.R. Anderson, Cosmic rays at balloon altitudes and the solar cycle, J. Geophys. Res. 67:1309.
- Nicolet, M. (1945): Contribution a l'etude de la structure de l'ionosphere. Mém.Inst. Royal Met. Belgique, 19:74.
- Nicolet, M. (1958): Aeronomic conditions in the mesosphere and lower thermosphere. Penn. State Univ. Ionosph. Res. Lab. Rept. 102
- Nicolet, M. (1978): private communication.
- Nicolet, M., and A.C. Aikin (1960): The formation of the D-region of the ionosphere. J. Geophys. Res., 65:1469.
- O'Brien, W.E. (1970): The prediction of solar proton events based on solar radio emissions. AFCRL-70-0425, Hanscom AFB, MA. 01731.
- Offermann, D. (1978): Recent advances in the study of D-region winter anomaly. J. Atmos. Terr. Phys., in press.
- Paulikas, G.A. (1975): Precipitation of particles at low and middle latitudes. Rev. Geophys. Space Phys., 13:709.
- Paulikas, G.A., J.B. Blake, and G.C. Freden (1968): Low-energy solar-cosmic ray cutoffs: Diurnal variations and pitch-angle distributions. J. Geophys. Res. 73:87.
- Paulsen, D.E., R.E. Huffman, and J.C. Larrabee (1972): Improved photoionization rates of O<sub>2</sub> (<sup>1</sup>Δg) in the D-region. Radio Sci., 7:51.
- Poppoff, I.G., R.C. Whitten, and R.S. Edmonds (1964): The role of nonflare X radiation in the D-region. J. Geophys. Res., 69:4081.



- Rees, M.H. (1969): Auroral electrons. Space Sci. Rev., 10:413.
- Reid, G.C. (1961): A study of the enhanced ionization produced by solar protons during a polar cap absorption event. J. Geophys. Res., 66:4071.
- Reagan, J.B. (1977): Ionization processes. In: Dynamical and Chemical Coupling Between the Neutral and Ionized Atmosphere. (B. Grandel and J.A. Holtet, Eds.) NATO Adv. Study Inst., D. Reidel Pub. Co., Dordrecht, Holland.
- Reagan, J.B., and T.M. Watt (1976): Simultaneous satellite and radar studies of the D-region ionosphere during the intense solar particle events of August 1972. J. Geophys. Res., 81:4579.
- Rusch, D.W., and C.A. Barth (1975): Satellite measurements of nitric oxide in the polar region. J. Geophys. Res., 80:3719.
- Sastri, J.H., and Murthy, B.S. (1975): Geomagnetic effect associated with x-ray flare from Sco X-1. Nature, 257:35.
- Shimabukuro, F.I. (1970): The observation of 3.3-mm bursts and their correlation with soft x-ray bursts. Solar Phys., 15:424.
- Smart, D.F., and M.A. Shea (1971): Proposed solar proton event classification system. Solar Phys., 16:484.
- Strobel, D.F., T.R. Young, R.R. Meier, T.P. Coffey, and A.W. Ali (1974): The nighttime ionosphere: E-region and lower F-region. J. Geophys. Res., 79:3171.
- Strickland, D.J., D.L. Book, T.P. Coffey, and J.A. Fedder (1976): Transport equation techniques for the deposition of auroral electrons. J. Geophys. Res., 81:2755.
- Svennesson, J., F. Reder, and J. Crouchley (1972): Effects of x-ray stars on VLF signal phase. J. Atmos. Terr. Phys., 34:49.
- Swider, W. (1964): The determination of the optical depth at large solar zenith distances. Planet. Space Sci., 12:761.
- Swider, W. (1969): Ionization rates due to the attenuation of 1-100A non-flare solar X rays in the terrestrial atmosphere. Rev. Geophys., 7:573.
- Swider, W. (1972): Reply, J. Geophys. Res., 77:2000.
- Swider, W. (1977): Aeronomic aspects of the polar D-region. Space Sci. Rev., 20:69.
- Swider, W. (1978): Daytime nitric oxide at the base of the thermosphere. J. Geophys. Res., 83:4407.

- Swider, W., and R.S. Narcisi (1977): Auroral E-region: Ion composition and nitric oxide. Planet. Space Sci., 25:103.
- Thorne, R.M. (1977): Influence of relativistic electron precipitation on the lower ionosphere and stratosphere. In: Dynamical and Chemical Coupling Between the Neutral and Ionized Atmosphere. NATO Adv. Study Inst., D. Reidel Pub. Co., Dordrecht, Holland.
- Triska, P., and J. Laštovička (1970): On the x-ray control of the ionospheric absorption of HF radio waves, Solar Phys., 15:504.
- Ulwick, J.C. (1973): Steady-state coefficients in the D-region during solar particle events. Space Res., 13:581.
- Velinov, P. (1970): Solar cosmic ray ionization in the low ionosphere. J. Atmos. Terr. Phys., 32:139.
- Velinov, P. (1974): Cosmic ray ionization rates in the planetary atmospheres, J. Atmos. Terr. Phys., 36:359.
- Whalen, J.A., J. Buchau, and R.A. Wagner (1971): Airborne ionospheric and optical measurements of noontime aurora, J. Atmos. Terr. Phys., 33:661.
- Williams, D.J. (1976): SELDADS; an operational real-time solar-terrestrial environment monitoring system. NOAA Tech. Rept. ERL 357-SEL 37, Supt. of Documents, U.S. Govt. Printing Office, Washington, D.C. 20402.
- Wratt, D.S. (1976): Ionization enhancement in the middle latitude D-region due to precipitating high energy electrons. J. Atmos. Terr. Phys., 38:511.
- Zmuda, A.J., and T.A. Potemra (1972): Bombardment of the polar-cap ionosphere by solar cosmic rays. Rev. Geophys. Space Phys., 10:981.

# EFFECTS OF D-REGION IONIZATION ON RADIO WAVE PROPAGATION

T R Larsen  
Norwegian Defence Research Establishment  
N-2007 KJELLER, Norway

The effects of anomalous D-region ionization upon radio wave propagation are described for the main types of disturbances: Sudden ionospheric disturbances, relativistic electron events, magnetic storms, auroral disturbances, polar cap events and stratospheric warmings. Examples of radio wave characteristics for such conditions are given for the frequencies between the extremely low (3-3000 Hz) and high (3-30 MHz) frequency domains. Statistics on the disturbance effects and radio wave data are given in order to contribute towards the evaluation of possibilities for predicting the radio effects.

## 1. INTRODUCTION

The aim of this review is to survey the various effects that D-region ionization has on radio wave propagation. In this context the ionospheric D-region is defined as that part of the earth's ionized atmosphere below 90 km which exercises influence upon the propagation characteristics. In practice this region extends down to 50-60 km, but may be as low as 40 km during the most severe disturbances in the polar areas. The so-called ionospheric C-layer around 65 km is thus included in our considerations.

It is assumed that only the abnormal ionization effects are of interest in this context. The normal D-region ionization and the propagation modes it sustains are therefore not considered here.

Except for brief mention of absorption of HF cosmic noise only propagation within the earth-ionosphere waveguide is treated. Transionospheric propagation of man-made signals falls outside the scope of this article. Furthermore, since only effects resulting from natural disturbances are of concern here, no references will be given to effects from the ionospheric heating experiments using powerful radio wave transmitters, nor of high-altitude nuclear explosions.

In Table 1.1 is listed the radio frequency domains, their propagation modes and primary uses which will be discussed (the table is after Davies (1978a), only extended to the ELF region).

A recent overview of the problem of forecasting the ionospheric parameters of interest to radio propagation is given by Davies (1978b). The AGARD Conference preprint from the Symposium on "Operational modelling of the aerospace propagation environment" (AGARD-CPP-238, 1978) contains valuable contri-



Table 1.1 THE RADIO SPECTRUM AFFECTED BY ANOMALOUS D-REGION IONIZATION

NAME	FREQUENCY RANGE	PRIMARY PROPAGATION MODES	PRIMARY USES
Extremely Low Frequency (ELF)	3-3000 Hz	Zero order mode (TEM) (Waveguide between ground and lower ionosphere)	Communication
Very Low Frequency (VLF)	3-30 kHz	Waveguide and ground wave	Navigation, Communications, Standard Frequency and Time
Low Frequency (LF)	30-300 kHz	Waveguide and Ground Wave	Maritime, Loran C, Broadcasting
Medium Frequency (MF)	300-3000 kHz	E-Region Reflection (Night) and Ground Wave	Maritime Aeronautical, International Distress, AM Broadcasting, Maritime and Land Mobile
High Frequency (HF)	3-30 MHz	Reflection from E- and F-Regions	Maritime and Aeronautical Fixed Services, Broadcasting, Amateurs, Citizens
Very High Frequency (VHF)	30-300 MHz	Line of Sight, Scatter from Ionospheric Irregularities, E <sub>s</sub>	Television, FM Broadcasting, Public Safety, Mobile, Aeronautical

butions. The proceedings edited by Agy (1970) on "Ionospheric forecasting" can also be recommended (AGARD-CP-49, 1970). Rawer (1975) has discussed the historical developments in relation to HF forecasting.

### 1.1 Variations in the reflecting and absorbing properties of the D-region

The most important parameter of the D-region in changing its reflecting and absorbing properties is the free electron density ( $N_e$ ) with its time- and spatial variations. The underlying causes for these changes may be very complex, e.g. due to variations in D-region composition or chemistry, or due to changes in the ionizing agents (Reagan 1977). (The reader is also referred to the papers in this issue by Arnold and Swider, respectively for a discussion of these phenomena.) The regular diurnal and seasonal variation of  $N_e$  will not be treated in depth here, but some considerations will be devoted to variations with the solar cycle.

The height regions between say 60 and 90 km are normally only lightly ionized: the electron density increases from around 10 to  $10^4$  electrons/cm<sup>3</sup>

over this height interval (Belrose 1969, Gregory & Manson 1970, Beynon & Williams 1976). Whereas the  $N_e$  variations at equator and out to  $40-45^\circ$  latitude are primarily under solar control, at greater latitudes irregular variations become more and more important. The data provided by Torkar (1977) underline this last point, cf Figure 1.1 which shows a collection of electron density profiles observed during 1968-76 using rocket experiments from the closely situated Andøya and Kiruna Rocket Ranges ( $\sim 69^\circ N$ ) in Northern Scandinavia (profiles F43/44 are for a low latitude station,  $37^\circ N$ ). Even though both day- and nighttime profiles are represented, the figure nevertheless illustrates the large ranges in  $N_e$  variations that are to be expected for a northern latitude location (see also e.g. Belikovitch et al 1975, Zelenkova 1976).

Similarly, there are spatial and temporal changes in the  $N_e$  concentrations also at lower latitudes. This can be inferred in a striking way from the satellite data of precipitating energetic electrons ( $>160$  keV) published by Imhof et al (1976), see Figure 1.2. These profiles of precipitating fluxes vs latitude illustrate both time spatial variations. Such electrons precipitation which will ionize the atmosphere below 75-80 km even equatorwards of the L-shell,  $L = 4$ , often dominate over other ionizing agents (see review by Paulikas 1975).

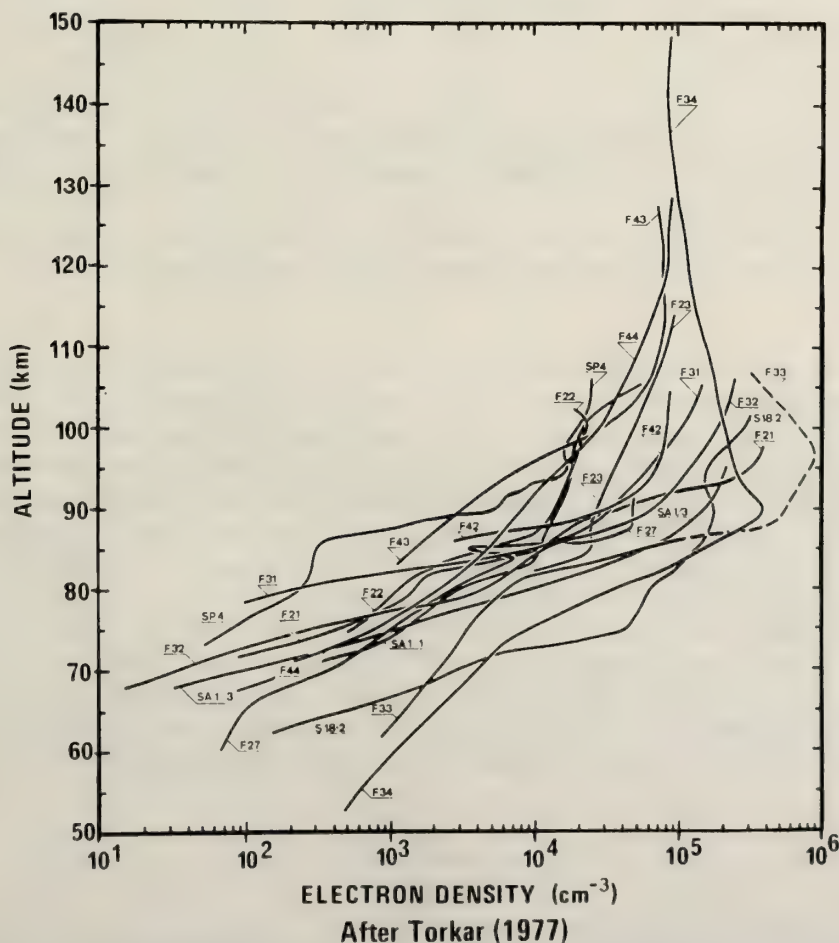


Figure 1.1 ELECTRON DENSITY PROFILES OBSERVED DURING 1968-76 USING ROCKET MEASUREMENTS (TORKAR 1977).

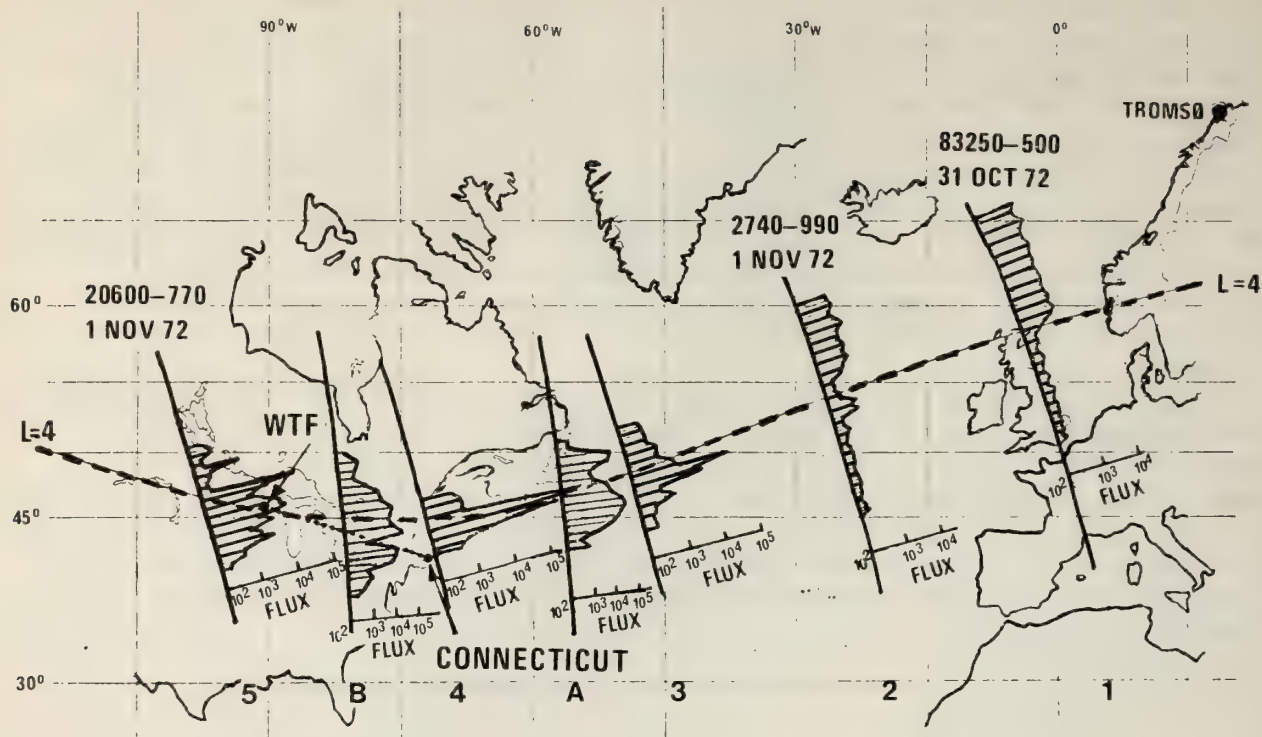


Figure 1.2 FLUXES OF PRECIPITATING ELECTRONS (>160 keV) AS MEASURED ON CONSEQUENTIVE PASSES BY TWO POLAR ORBITING SATELLITES (IMHOF ET AL 1976). PRECIPITATING FLUXES ARE HIGHER OVER THE US THAN OVER EUROPE. PASSES 4 AND 5 ARE SIMULTANEOUS WITH PASSES A AND B, RESPECTIVELY. BOTH TIME AND SPACE VARIATIONS ARE INFERRED FROM THESE AND OTHER SIMILAR DATA.

From such data one must conclude that the ionosphere exhibits significant time- and spatial variation in the  $N_e$  concentrations at all D-region heights. It is probably realistic to also conclude that a comprehensive global and real-time monitoring system of such variations having the necessary capability to "report" the detailed complexities of the ionized atmosphere will not be forthcoming. In the foreseeable future less than optimum solutions will be sought, providing real time coverage of limited areas, statistical models, or forecasts to operators of radio wave systems.

The D-region reflects VLF and LF waves from levels around 65-75 km at daytime and 80-90 km during night. For higher frequencies, above say 1 MHz, which are reflected from the E- or F-layers, the D-region will act mainly as an absorbing medium. For the principal polarization at these frequencies ( $f$ ) the absorption in the case of quasi-longitudinal propagation is proportional to:

$$\frac{N_e \nu}{(f + f_L)^2}$$

where  $f_L$  is the longitudinal component of the gyrofrequency, and  $\nu$  the collision frequency. At lower frequencies ( $2\pi(f + f_L) \ll \nu$ ) absorption is expected to vary like:  $N_e/\nu$ .



At ELF frequencies, especially during nighttime, an appreciable contribution to the wave field comes from levels above the D-region (Booker & Lefeuvre 1977, Barr 1977, Imhof et al 1976, 1977).

There are several approaches to the problem of deducing the electron density profile from radio propagation data (cf. eg.: Bain & May 1967), but this is a complicated subject which will not be discussed.

## 2. SUDDEN IONOSPHERIC DISTURBANCES (SIDs)

In the context of this article the interesting effects of solar flares are the solar X-rays (SXR) and the extreme ultraviolet (EUV) emissions which cause disturbances in the lower ionosphere at heights above 60-70 km on the sunlit hemisphere. Emission of energetic particles from flares will be discussed in section 5. The X-ray fluxes in the 1-8 Å band can increase by several orders of magnitude, whereas the EUV variation is less than a few per cent during a flare. The emissions cause enhanced ionization at D-region heights and may change the ion chemistry (Jones 1967, Rowe et al 1970, Montbriand & Belrose 1972, Desphande & Mitra 1972, Laštovička & Smilauer 1973, Mitra 1974, Belikovitch et al 1977). The direct ionization rates depend critically upon the X-ray intensities and their spectral distribution (see article by Swider in this volume on production rates). An example of changes in the free electron content at D-region heights during a flare is illustrated in Figure 2.1 (Rowe et al 1970). Note that the ionization changes are different both in magnitude and time history at the different levels.

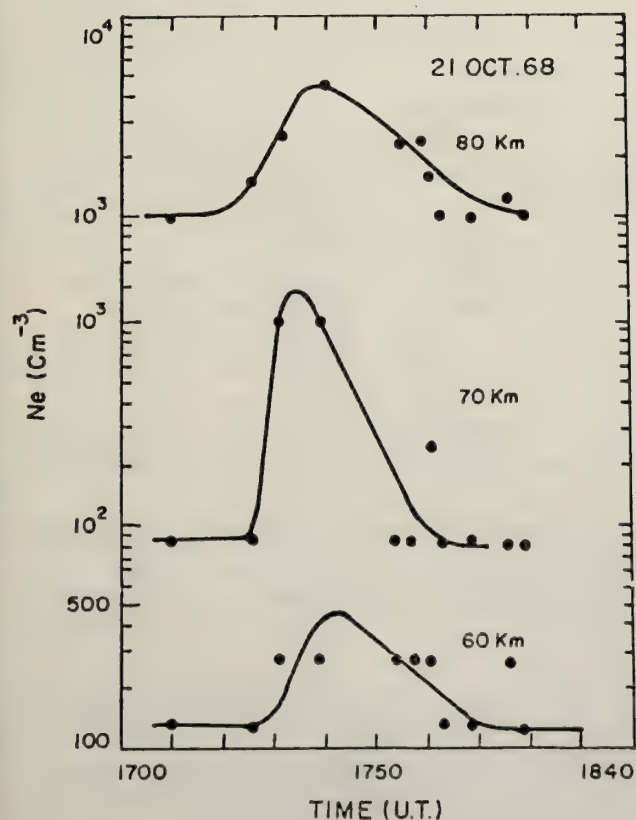


Figure 2.1 ELECTRON DENSITY VARIATIONS AT THREE D-REGION HEIGHTS OBSERVED BY THE WAVE INTERACTION TECHNIQUE DURING THE SID OF OCTOBER 21, 1968 (ROWE ET AL 1970).

New satellites with enhanced capabilities to measure the detailed energy spectra of solar X-rays have greatly increased the possibilities of predicting the direct effects upon the atmosphere, and in almost real time (Donnelly 1976, Rothmuller 1978). For accurate prediction of the effects on the radio wave propagation better knowledge of the D-region ion chemistry and the electron loss rates is necessary, as well as more accurate information on the neutral atmosphere (e.g. density, temperature and composition). At low heights, <70 km, Thrane (1978) concludes that only the gross effects of X-ray flares can be predicted due to uncertainties in the ion chemistry.

The effects of solar X-rays and EUV in the atmosphere are generally denoted by the term sudden ionospheric disturbances (SIDs). The monograph by Mitra (1974) on "Ionospheric Effects of Solar Flares" gives a thorough treatment of SIDs, and is highly recommended for an in-depth discussion of these phenomena. The book also provides a valuable bibliography for this purpose.

There are several aspects of SIDs and Table 2.1 lists the more important effects from a radiopropagation point of view. The approximate frequency bands effected are also listed, as are the recommended acronyms (Mitra 1974) for the disturbances.

Table 2.1 EXAMPLES OF DIFFERENT ASPECTS OF SIDs

PHENOMENON	FREQUENCY BANDS INVOLVED
Sudden Decrease of Atmospherics	(SDA) ELF, VLF, LF
Sudden Enhancement of Atmospherics	(SEA)
Sudden Field Anomaly	(SFA) VLF, LF
Sudden Phase Anomaly	(SPA) VLF, LF
Sudden Frequency Deviation	(SFD) HF
Short-wave Fadeout	(SWF) HF, Lower VHF
Sudden Cosmic Noise Absorption	(SCNA) HF, Lower VHF
Sudden Increase in Total Electron Content	(SITEC) VHF

Most of the increase in total electron content during a flare takes place above the D-region and this phenomenon is therefore not considered further. The sudden frequency deviations at HF also occur at heights above ~100 km (Donnelly 1968).

## 2.1 Sudden increase/decrease of atmospherics

It has been known for many decades that the field intensity of atmospherics changes during a solar flare. Depending upon the investigating frequency increases or decreases in intensity may be observed. Enhancement of atmospherics (SEA) occurs approximately between 10 and 75 kHz, the latter frequency being dependent upon angle of incidence (Sao et al 1970, Mitra 1974) of the waves. At ELF frequencies, below ~1 kHz enhancement also occurs, whereas between 1 and 10 kHz a signal decrease (SDA) is expected and observed (Sao et al 1970, Triska & Lastovicka 1972, Lastovicka 1977). Results for ~300 SEAs are reported by Leroy (1978, this issue).

## 2.2 Sudden field and phase anomalies

When one discusses variations of stable man-made VLF and LF radio transmissions during a flare, changes in amplitude are denoted by SFA (sudden field anomaly) whereas sudden phase anomalies are called SPAs. These phenomena are discussed by e.g. Chilton et al (1965), Burgess & Jones (1967), Yamashita (1969), Wakai et al (1973), Sengupta (1974), and Kaufmann & Piazza (1978).

The detailed effects of SIDs upon VLF and LF phase and amplitude depend upon a number of parameters in addition to the time- and energy characteristics of the SXR emission itself and the D-region chemistry: frequency of VLF waves, path characteristics and propagation distance, solar illumination over the path and season. For specific paths and wave frequencies, however, estimation of the average SXR induced variations should be possible. With real-time modelling of the solar X-ray fluxes (Rothmuller 1978, Argo & Rothmuller 1978, this issue) even better results should be potentially available. Figure 2.2 illustrate, however, the complexities of predicting the various SID effects (Rose et al 1971). The phase and amplitude changes at five frequencies in the 21-32 kHz band over two paths of slightly different path lengths along a common great circle trajectory are intricate and exhibit presence of multi-mode propagation effects. Note e.g. the amplitude variation at 31 kHz - the changes

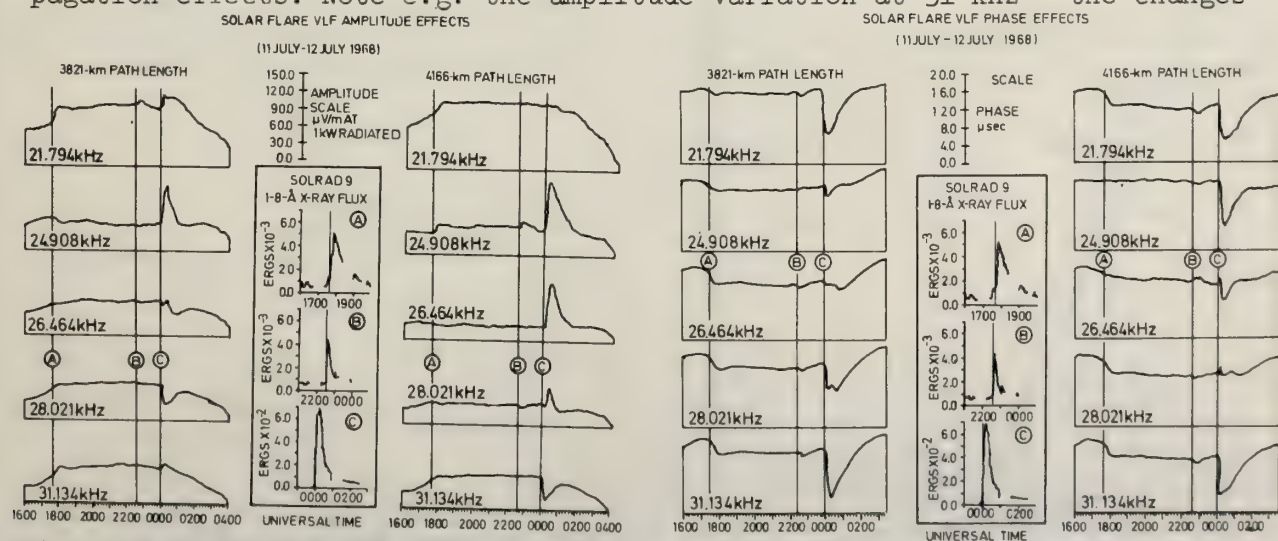


Figure 2.2 SOLAR FLARE EFFECTS OBSERVED ON 11-12 JULY 1968 ON THE AMPLITUDE AND PHASE AT 5 FREQUENCIES BETWEEN 21-32 kHz FOR TWO TRANSMISSION PATHS (ROSE ET AL 1971).



have opposite sense due to mode mixing. Cf. also Chakravarty & Ramanatan (1972) and Apostolov (1975).

Detailed information on the SXR emissions are required to model the radio anomalies, as a general result, however, Bain and Hammond (1975) found that fluxes greater than  $6 \cdot 10^{-7} \text{ J m}^{-2} \text{ s}^{-1}$  in the 0.5-3.0 Å band always produced a SPA on VLF.

### 2.2.1 Influences of SFAs and SPAs on operating navigation system.

Several studies have been devoted to effects of solar X-rays on Omega navigation frequencies (10-14 kHz), cf. review by Swanson (1977). In Figure 2.3 is shown the probability distribution for phase changes at 10.2 kHz on a long distance path for the years 1966-70 (Swanson & Kugel 1973), compared with the typical ("undisturbed") phase deviation distribution, having a standard deviation of 3 centicycles (cec). (The probability values for flare conditions should be reduced by a factor of 4 to correspond to normal, 24 hours daily operation.)

Swanson and Kugel furthermore found from the analysis of 500 events observed over two midlatitude paths (Hawaii and Trinidad to New York) a typical duration of 40 min for the SPA, with a risetime to maximum phase offset of approximately 6 min. Mean maximum phase offset was observed to be 23 cec.

Larsen (1977) analysed Omega phase difference data at 10.2 kHz for two sets of station pairs in 1968: Omega Norway minus Trinidad recorded at Oslo (60°N) and Omega Norway minus Hawaii recorded at Svalbard (78°N). The mean maximum phase offset from expected values was 23 cec (corresponding to 3-4 km error if the disturbances only affected one set of the hyperbolic lines of position (LOPs) - two LOPs are needed for a position fix). The

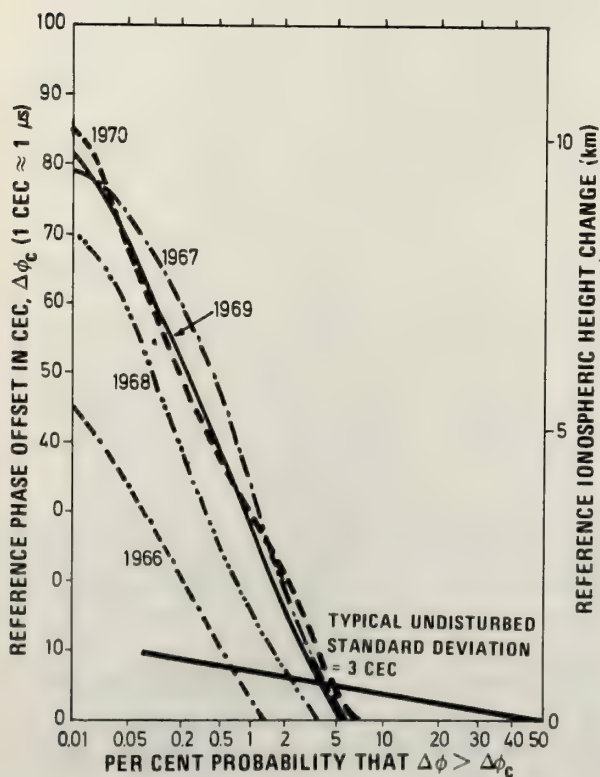


Figure 2.3 PER CENT PROBABILITY OF DISTURBANCE LEVEL OCCURRENCE FOR OBSERVED SPA EVENTS ON THE HAWAII TO NEW YORK PATH AT 10.2 kHz (SWANSON & KUGEL 1973).

largest position fix error was about 10 km, which appeared on 8 July 1968 1730 UT (cf. also Swanson 1971). Average duration of the SPAs was about 55 min. In 1968 near the maximum of the solar cycle the total duration of conditions with SPA effects was less than 1% (Larsen 1977). See also Burgess and Walker (1970), Penny (1977) and Jones (1978a).

On Loran-C frequencies SFA effects have also been detected (Jean & Large 1967). Wakai et al (1973) reports on Loran-C measurements in Japan over path-lengths up to about 3000 km.

Measurements such as the ones reported by Jean and Large (1967) will not automatically lead to deteriorations in navigational accuracy. In areas where the ground wave of the Loran-C pulses are used for navigation, conditions should essentially be unaffected by SIDs unless the increased strength of the Loran-C skywave interferes with the groundwave reception. Possible problems with skywave arrival before the receiver sampling time of the groundwave will be briefly discussed in section 5. Belrose (1968) has described polarization changes in the skywave at Decca frequencies during a flare.

### 2.3 Short-wave fadeout and sudden cosmic noise absorption

Short-wave fadeout (SWF), or the Møgel-Dellinger effect are decreases or complete fadeouts of HF signals during the visible flares. Examples of the three characteristic types: sudden, slow and gradual are shown in Figure 2.4 (Mitra 1974). Effects are also seen at lower frequencies (LF and MF) and at VHF frequencies enhancements may be observed, due to enhancement of scatter reflections.

Sudden cosmic noise absorption (SCNA) is a decrease in cosmic radio noise received by riometers. Since the discovery of the effect by Shain and Mitra (1954), it has been studied by a large number of groups and the riometer has been a valuable equipment both in case studies and synoptic studies of widespread absorption. Using data from several riometer frequencies it is possible to unfold the electron density profile as a function of height (Belikovitch et al 1971, Rangasawamy 1974).

Riometer absorption is further discussed in sections 4 and 5, see also Lavergnat et al (1976) for a critical discussion of application of riometer measurements in ionospheric studies.

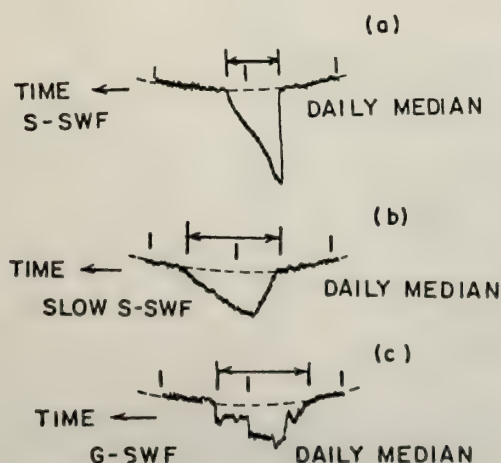


Figure 2.4 REPRESENTATIVE SHAPES OF DIFFERENT TYPES OF SHORT WAVE FADEOUTS (SWFs). CURVE (a) SHOWS A TYPICAL S-SWF; CURVE (b), A SLOW S-SWF; CURVE (c), A G-SWF (MITRA 1974).

### 3. RELATIVISTIC ELECTRON PRECIPITATION EVENTS

Bailey and Pomerantz (1965) concluded from a study of cases with anomalous daytime absorption on HF forward scatter links in Alaska that the absorption was caused by unusual ionization below the scattering stratum of the waves. Electrons with energies common in auroral displays (1-10 keV), were insufficient ionizers since no absorption events were detected during times of auroral activity. To penetrate the atmosphere below the assumed HF scattering heights (~70 km) particle energies of  $\geq 400$  keV and  $>10$  MeV were necessary for electrons and protons, respectively. Only electrons are trapped in the geomagnetic field with sufficient fluxes and energies at the L-values where the anomalous absorption most frequently occur to cause the anomalous ionization, hence the term "relativistic electron precipitation" (REP) events are used to identify such disturbances (see Figure 3.1 for examples of REP energy spectra, Thorne 1978, private communication).

The REP events occur mainly between  $L = 4.5$  and 6 (Bailey & Pomerantz 1965). No events were detected by these authors on HF forward scatter paths having their mid-points beyond  $L = 10$  (see also Larsen & Thomas 1974, Imhof et al 1978). The events seem to have a large longitudinal extent. Judging from a study of riometer records for a REP event in May 1972, Larsen (1974) concluded that the extent for that case was at least 2000 km.

The REP events, as found by Bailey (1968) occur most frequently in the time period 06 to 18 hours local time. Most events last from 1 to 6 hours, but there are cases where one event seems to extend from one day to the next. The histogram in Figure 3.2 shows the monthly number of REP occurrences with absorption  $\geq 8$  dB for the years 1964 to 1966 (Bailey 1968) and the first 6 months of 1969 (Bailey, private communication, 1971). Note the tendency

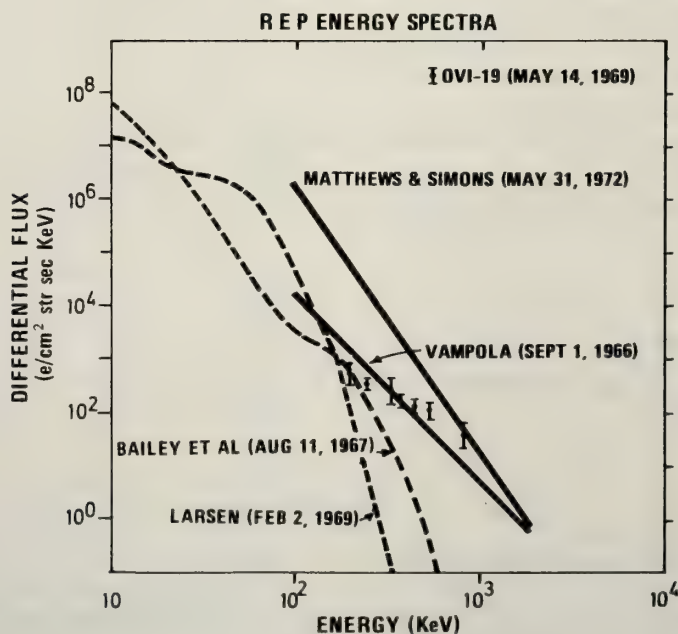


Figure 3.1 COLLECTION OF ENERGY SPECTRA OF PRECIPITATING ELECTRONS DURING DIFFERENT REP EVENTS (THORNE 1978, PRIVATE COMMUNICATION).



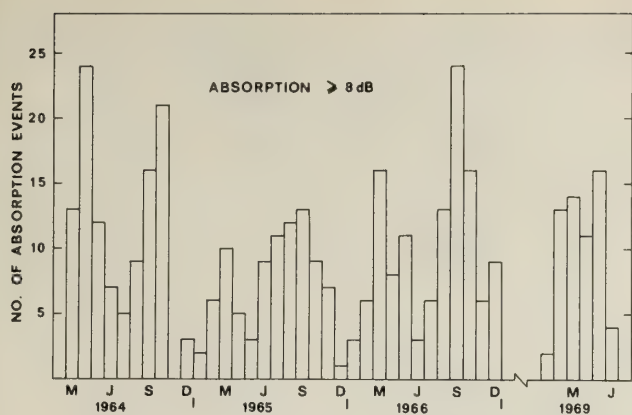


Figure 3.2 HISTOGRAM OF FORWARD SCATTER ABSORPTION EVENTS FOR THE YEARS 1964-1966 (BAILEY 1968) AND FIRST 6 MONTHS OF 1969 (BAILEY, PRIVATE COMMUNICATION 1971).

has studied the effect of D-region absorption on the forward-scatter radio meteor amplitude distribution index.

A number of studies have indicated that REP events are one aspect of magnetospheric substorms even though there are more substorms than REP events (Rosenberg et al 1972).

Other studies (Larsen & Thomas 1974, Matthews & Simon 1973, Thorne & Larsen 1976, Imhof et al 1978) provide further information on the disturbances and the energy spectra of the electrons causing the excess ionization.

Apart from Bailey's studies there does not seem to have been systematic investigations of the duration, frequency of occurrence, latitude and longitude coverage of the REP disturbances with the aim of assessing their influence on radiowave propagation. At present, we do not understand the physical mechanisms behind the REPs well enough to predict such events.

for a larger number of REP events near the equinoxes approaching 25 cases per month. The local magnetic disturbance during an REP is often small and even though they have no close connection to the SC of storms, they frequently occur during geomagnetically disturbed periods (Bailey 1968).

Abnormal nighttime VLF effects have furthermore been detected on a short propagation path in North-Norway in a number of cases when daytime REP activity was detected over Alaska (Larsen 1973, Thorne & Larsen 1976). Belrose (1968) has provided numerous cases of propagation anomalies due to high energy particles. Jones and Collins (1973)

#### 4. MAGNETIC STORMS AND AURORAL DISTURBANCES

##### 4.1 Radio wave absorption

It is well established that significant changes in the ionization of the mid-latitude ( $L \approx 2-4$ ) D-region may occur during and particularly following magnetic storms and disturbances to mid-latitude LF and VLF radio transmissions associated with storms have been observed for a half century (Bracewell et al 1951, Lauter 1961, Lauter and Knuth 1967, Belrose and Thomas 1968). Studies of this phenomenon have described the effects in terms of "primary storm" and "after storm" effects which occur during the main phase and the recovery phase, respectively, of the geomagnetic storm. The term "poststorm" effect has been used by Montbriand and Belrose (1976) to describe excess ionization in the D-region persisting after the magnetic disturbance has subsided. Several authors have suggested that precipitating energetic electrons ( $E \geq 40$  keV) produce the enhanced ionization in the D-region that causes these disturbances (e.g., Lauter & Knuth 1967, Belrose & Thomas 1968).

Observations of energetic electron precipitation at middle latitudes have been made with rocket and satellite instruments (e.g., Gough & Collin 1973, Aiking 1974, Larsen & Thomas 1974, Potemra 1973) which predict that these particles are a significant ionization source in the lower ionosphere. For the magnetic storm of 16 December 1971, Larsen et al (1976) found that precipitating electrons from the earth's radiation belt were the main cause of the D-region poststorm effects at middle latitudes. A theoretical study by Spjeldvik and Thorne (1975) shows that radiation belt electrons resonating with naturally occurring whistler mode waves can adequately account for delayed ionization enhancement in the D-region following geomagnetic storms. But there is not complete agreement as to the cause of the after storm effect, as changes in the ionospheric chemistry cannot generally be ruled out as a possible cause (see Belrose and Thomas (1968) for a discussion). Paulikas (1975) has reviewed particle precipitation in middle latitudes.

Aspects of absorption of radio waves in the lower ionosphere has in recent years been studied by Driatskiy et al 1971, Komrakov et al 1972, Cipriano et al 1974, Laštovička 1975, 1976.

#### 4.1.1 Propagation at VLF and LF in middle latitudes during and following magnetic storms

Potemra and Rosenberg (1973) have presented phase disturbances to long-distance mid-latitude ( $2 \leq L \leq 4$ ) VLF transmissions and simultaneous observations of bremsstrahlung X-rays near  $L = 4$  during a magnetospheric substorm as evidence for ionization enhancements due to electron precipitation. By correlating these data with broad band VLF emission data these same authors concluded that the electron precipitation was caused by a cyclotron resonance between energetic trapped electrons and whistler mode VLF waves. Larsen et al (1976) have combined observations of precipitating electrons ( $>130$  keV) obtained from a polar-orbiting spacecraft with electron density profiles measured with a ground-based partial reflection technique at Ottawa during several geomagnetic storms to confirm that D-region ionization enhancements are caused by precipitating energetic electrons at these latitudes ( $L = 3.65$ ). Long-lasting effects (lasting several days) at VLF have been detected (Larsen et al 1977). Furthermore Belrose (1968) and Reder & Westerlund (1970) have provided numerous examples of erratic VLF and LF phase and amplitude variations during magnetic storms. Singer & Bremer (1974) have discussed D-region electron density profiles for various solar activity levels in relation to propagation at LF, MF and HF, cf. also Bleiweiss et al (1973).

Comprehensive studies of the behaviour of LF waves over an 18 year span has been reported by Lauter and Sprenger (1952) and Lauter and Knuth (1967). They found that:

- The sudden commencement (SC) of a magnetic storm was not followed by any change in signal strength
- During nighttime, rapid and deep fading accompanied the main phase of the storm ("primary storm effect")
- Daytime observations showed no effects during the main phase of the storm
- Following certain storms the nighttime signal strength was weak for a prolonged period starting 3-4 days after the start of the storm (storm "after effects")



Nestorov (1977) has described signal variations of LW waves (164 kHz) over a 17 year period and found that the long term variations of the LF field:

- Correlate negatively with the intensity of cosmic rays, and
- Positively with sunspot numbers.

For night conditions he infers that there exists a zone in the mesosphere in which the ionization density is determined by two sources which vary in opposite sense with the sunspot cycle.

#### 4.1.2 Propagation at ELF

Since 1970 a number of ELF propagation studies have been made using the transmissions from US Navy Wisconsin Test Facility (WTF) in the USA. The signal frequencies used lie in the 40-50 and 70-80 Hz ranges. An extensive treatment of this subject is given in the Special Issue on Extremely Low Frequency Communications (IEEE Transactions on Communications, Vol. com-22, No. 4, 1974). The articles in that issue give a comprehensive bibliography for the ELF field.

Anomalies in the received signals have been detected for both mid-latitude and high-latitude paths (Bannister 1975, Davis 1976). Decreases in signal amplitude of more than 6 dB are occasionally observed for several middle latitude paths in the early hours local time (01-03 LT). The causes are believed to be of ionospheric origin, but at present no explanation seem to be completely satisfactory. The effect is probably not one of only signal absorption, but changes in the excitation of the waves into the waveguide must also be invoked (Bannister 1975, Imhof et al 1976, 1977, Booker & Lefeuvre 1977, Booker 1977, Barr 1977).

#### 4.2 The mid-latitude "winter anomaly"

Distinct from the poststorm effects is the D-region "winter anomaly", which is defined in terms of the increased and highly irregular day-to-day variations of absorption of HF waves reflected from the ionosphere at middle latitudes during winter conditions (Dieminger 1952, Appleton & Piggott 1954, Thomas 1962, 1968, Belrose 1969, Manson & Merry 1970, Schwentek 1971, Hargreaves 1973, Kalinovskaia & Anischenko 1973, Beynon et al 1976, Elling & Schwentek 1977, Hidalgo 1977).

Radio wave absorption in winter (between latitudes of approximately  $35^{\circ}$  and  $60^{\circ}$ ) does not follow the simple solar zenith angle dependence which is obtained for summer observations. This subject has recently been extensively investigated during the Western Europe Winter Anomaly Campaign 1975-76 (Offerman 1977, Rose & Widdel 1977, Thrane et al 1979).

The latter authors concluded that the enhanced absorption was due to increased electron densities in the height region 75-90 km. The studies revealed that this enhanced ionization was not due to increased ionizing radiation, but due to at least three factors:

- Enhanced density of mesospheric nitric oxide, NO, causing increased ionization by solar H - Lyman- $\alpha$  radiation.
- Decreased electron loss rate at heights of 75-85 km. This decrease coincided with a depletion of heavy positive water cluster ions with large recombination rates.



- Enhanced density of mesospheric excited molecular oxygen  $O_2$  ( $'\Delta_g$ ) causing increased ionization by solar UV radiation.

The period of enhanced absorption in the above studies (January 1976) was associated with wavelike structures which resulted in marked deviation of the atmosphere from the reference atmosphere. The disturbances could be traced to ground level and indicate the importance of ionosphere - atmosphere coupling.

#### 4.3 Absorption at low latitudes

Significant activity has been devoted towards explaining absorption at low latitudes (see e.g. Oyinloye 1975). D-region absorption at Dehli has been studied by Samra & Samra (1972) and Seera et al (1973). Effects have also been investigated by Ganguly (1972, 1975).

#### 4.4 Radio propagation during ionospheric substorms and auroral absorption

Several studies have been concerned with radio wave propagation during ionospheric substorms (Bates & Albee 1965, Reder et al 1964, Belrose 1968, Egeland et al 1969, Westerlund & Reder 1973, Svennesson 1973 and Leonov 1975).

The propagation effects are usually observed between, say  $65^\circ$  and  $70^\circ$  geomagnetic latitudes, but this range depends upon the substorm activity. At high activity levels an equatorwards movement is expected. Berkey et al (1974) have made a synoptic investigation of the precipitation patterns during 60 substorms during IQSY and IASY, cf. also Hargreaves & Cowley (1967) and Theander (1972).

Samuel and Bradley (1975) have presented a new form of representation of the diurnal and solar cycle variations of ionospheric absorption.

Svennesson (1973) has studied effects on VLF propagation during substorms. He investigated 16 events during 1969 using 8 receiver sites monitoring 9 transmitters. For all these events there existed synoptic riometer maps at frequent time intervals (part of an international cooperative investigation, Berkey et al 1971).

His findings are summarized as follows:

- All the observed VLF phase anomalies during the analysed ionospheric substorms could be explained by a depressed ionosphere due to enhanced particle precipitation, no mode conversion effects were observed.
- A good correlation was found between VLF phase anomalies and the ionospheric substorm as observed by riometers in both the day and night ionospheres. Figure 4.1 shows an example of the VLF and the absorption maps (refer to the paper for detailed discussion).
- The amplitudes of the received VLF signals showed in most cases no significant divergence from the normal values during the substorm. The amplitude was low by 0-4 dB, and the signal was noisier than usual during a few of the most severe VLF phase anomalies. This observation of no or very little decrease in the amplitude is consistent with the theoretical model of Wait and Spies (1964) which, for a constant ionospheric conductivity gradient, gives only a very small increase in attenuation of the first order mode due to a decrease in the reflection height.

It is interesting to note that strong effects on VLF signals propagating in the daytime hemisphere were observed during all the analysed events. This was interpreted as being due to energetic electrons ( $>200$  keV).

# 15 JANUARY 1969 EVENT

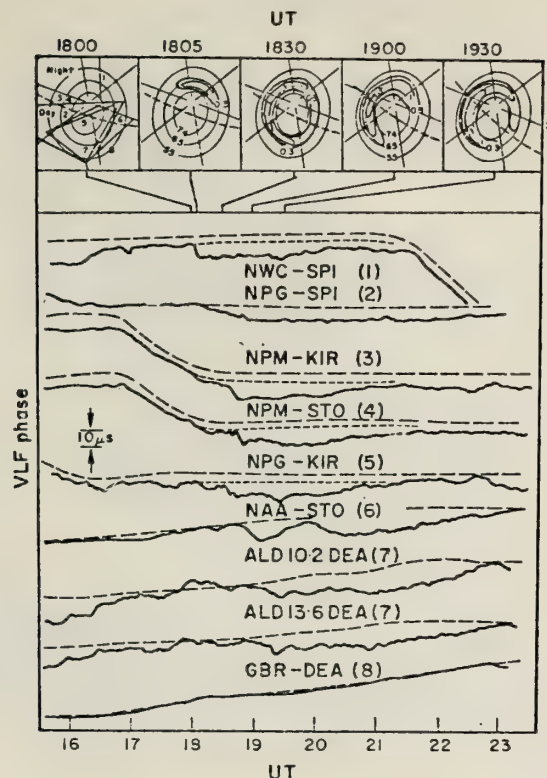


Figure 4.1 COLLECTION OF VLF PHASE RECORDS FOR AN IONOSPHERIC STORM ON 15 JANUARY 1968 (SVENNESSON 1973).

THE TOP ROW OF MAPS INDICATE THE REGION OF RIOMETER ABSORPTION.  
(REFER TO THE ORIGINAL PAPER FOR A COMPLETE DISCUSSION OF THESE DATA.)

Svennesson (1973) claims that the lack of detailed knowledge of the precipitation of high energy electrons along the path seems to be the major reason for the poor correlation so often reported between VLF effects and ionospheric disturbances.

Auroral phenomena are often accompanied by HF radio "blackout" periods. The cause for such excess ionization is twofold: electrons precipitating into the atmosphere in connection with the aurora, and energetic electrons having higher energies than the typical 1-10 keV electrons which cause the visual auroras. Hartz and Brice (1967) showed that the high energy electrons generally precipitated equatorwards of the auroral oval in an almost circular region, around geomagnetic latitudes of 63-67°. Figure 4.2 shows a statistical picture of the resulting HF riometer absorption (30 MHz) as a function of geomagnetic latitude and time (Hartz et al 1963). Ranta and Ranta (1977) have discussed the daily variation of D-region absorption at high latitudes.

The effects on radiowave propagation are also marked. Ordinary HF communication in polar areas is often impossible during "blackout" conditions. Figure 4.3 shows the occurrence of blackouts for a path between Svalbard (78°N) to North-Norway (68°N) (Veastad 1968) for the years 1956-67.

The results by Folkestad (1968a) in Figure 4.4 shows the lowest and highest observed frequencies (LOF and MOF) over two paths between North-Nor-



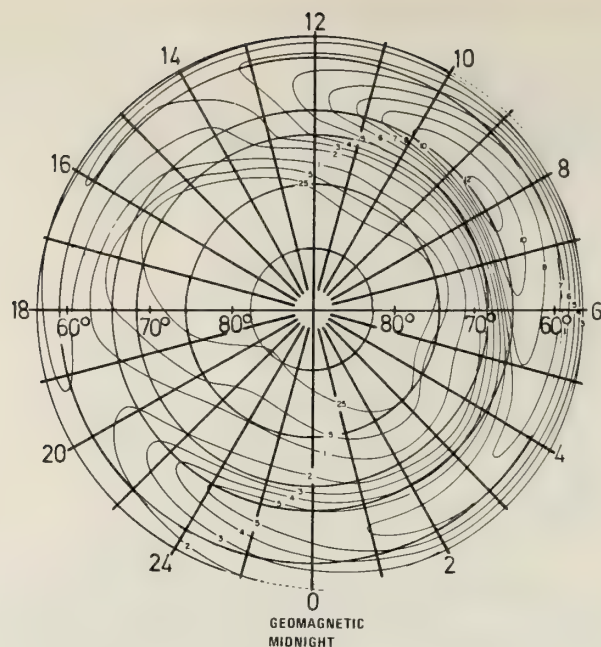


Figure 4.2 THE FIGURE SHOWS THE PERCENTAGE OF THE TIME THAT AURORAL RADIO WAVE ABSORPTION WAS 1.0 dB OR MORE AT 30 MHz (HARTZ ET AL 1963). DATA REFER TO THE NORTHERN HEMISPHERE AND ARE PLOTTED AS A FUNCTION OF GEOMAGNETIC LATITUDE AND MEAN GEOMAGNETIC TIME.

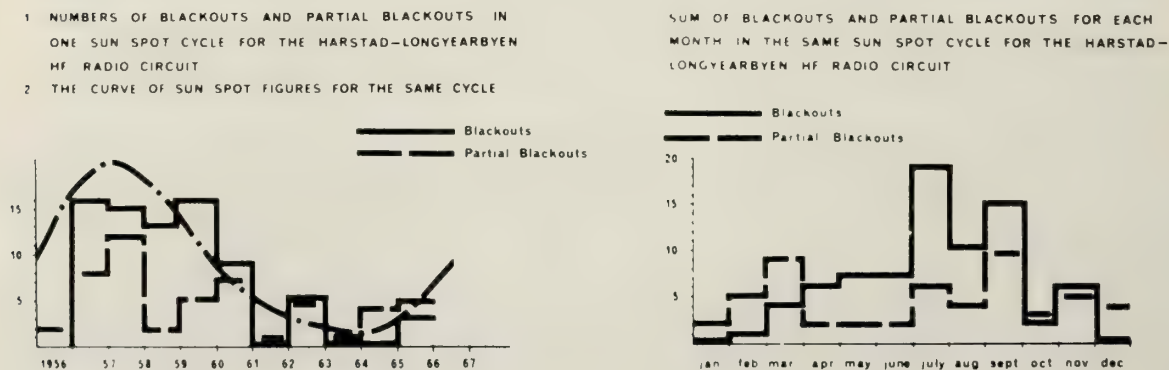


Figure 4.3 OCCURRENCE OF BLACKOUTS DURING SUNSPOT PERIOD (VEASTAD 1968).

way and USA. Periods of complete blackouts can be noted during the night hours. Barclay (1978) has discussed the statistical modelling of HF links. New probing techniques (Reilly 1977) may prove to be valuable.

Jones (1978b) has reviewed HF communication and possibilities for channel evaluation. Actual data are shown in Figure 4.5 which give sample results of systematic experiments evaluating the quality of HF communication at specific frequencies over two paths in North-Norway (Thrane 1978, private communication). The probability of establishing a connection around local midnight for summer conditions is plotted in % vs the probing frequency. Also plotted is the predicted maximum usable frequency (MUF) for the two paths. The signal



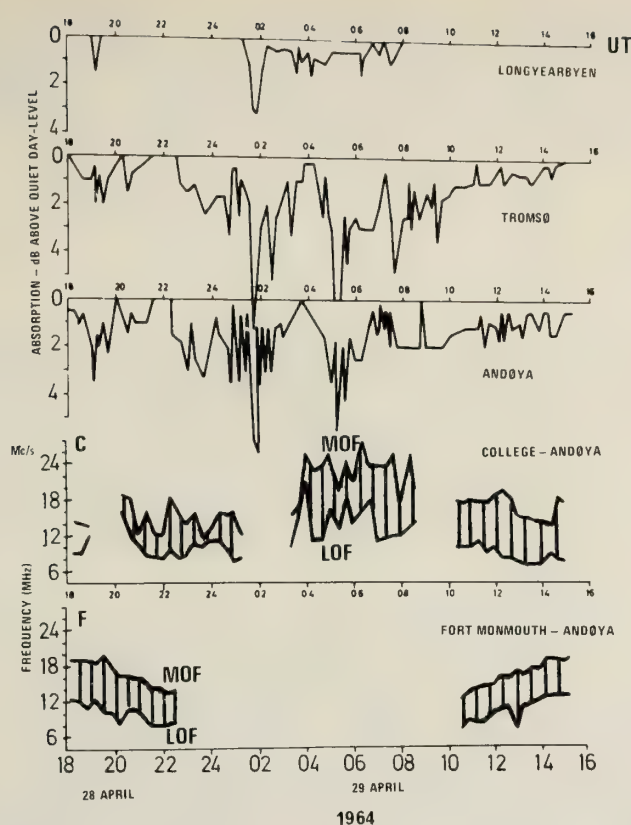


Figure 4.4 CIRCUIT BEHAVIOUR DURING DISTURBED CONDITIONS (FOLKESTAD 1968a).

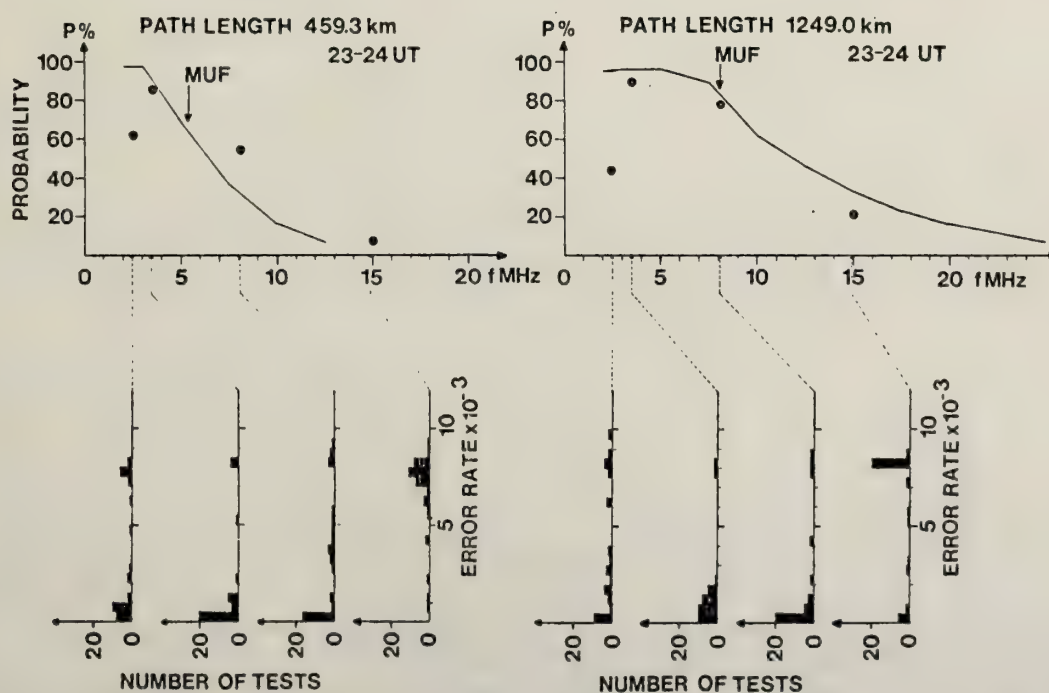


Figure 4.5 RESULTS FROM SYSTEMATIC EVALUATING THE QUALITY OF DIGITAL HF COMMUNICATION FOR TWO PATHS IN NORTH-NORWAY (THRANE 1978, PRIVATE COMMUNICATION). FOR EXPLANATION, SEE THE TEXT.

transmitted is in digital form and connection, for these purposes, is defined as being established when more than 90% of the bits are properly received. Quite good agreement is obtained between predictions and measurements at the 2-3 highest frequencies. At the lowest frequency (2.5 MHz) reception seems to be limited by low signal to noise ratio.

## 5. POLAR CAP EVENTS

Following certain solar flares energetic particles impinge upon the Earth's atmosphere and cause severe and long lasting effects in the polar atmospheres. The particle influx consists mainly of protons of energies up to 100 MeV, but heavier ions and electrons may also be present (cf. the topical review by Zmuda and Potemra (1972), and articles by Potemra and Zmuda (1972), Lanzerotti and MacLennan (1971), Lanzerotti (1972), Vampola (1969), Reagan (1975)). The energetic protons ionize down at low altitudes (down to ~30 km, Reagan (1975)) over a large spatial area, from the poles to 55-60° geomagnetic latitude. The proton fluxes are quite uniform over these areas, but asymmetries exist, mainly during the initial stages of the events over the central polar areas and at the low latitude cut-off regions (Evans & Stone 1969, Hakura 1970, Van Allen et al 1971, Blake & Vampola 1974). The influx in this latter region is time and energy dependent, especially at lower energies  $\leq 20$  MeV (Reagan et al 1972). The events are called polar cap events (PCE) or solar proton events (SPE). Historically the acronym PCA (for polar cap absorption) has frequently also been used for the events themselves.

The production of ion pairs (see review by Swider, this volume) during a PCE may increase by several orders of magnitude. For the most intense SPE observed by modern techniques (4 Aug 72) the peak production of ion pairs occurred at 40 km's altitude with a rate of  $3 \cdot 10^4 \text{ cm}^{-3} \text{ s}^{-1}$  (Reagan 1975) almost four orders of magnitude above nominal production at these levels. During such conditions D-region chemistry may actually become less complicated than during undisturbed conditions and the effective electron loss rates decrease (Reagan 1975). See also Zelenkova et al (1976).

Another SPE which has been extensively studied, using satellite, rocket- and ground-based observatories, is the November 1969 event, for which Ulwick (1969) has edited a collection of 49 papers describing the energetic particles and their ionospheric and D-region effects.

The frequency of occurrence of SPEs is correlated with the 11 year solar sunspot cycle. In Figure 5.1 is shown a comparison between the annual mean sunspot number ( $R_z$ ) and such events over nearly two solar cycles (1954-1963). The solar proton events (referred to as PCAs) are plotted vs time with the solar sunspot number at the time of the event (Pomerantz & Duggal 1974) as ordinate. It is clearly seen that more events are observed near the solar maximum periods than during quiet solar conditions, individual SPE events also tend to occur most frequently on days of high solar activity.

The most comprehensive list of SPEs available is published in the book edited by Švestka and Simon (1975). A total of 352 cases for the years 1955 to 1969 are documented as "confirmed" events.

Solar proton events may be single events or they can occur in overlapping sequences (cf. reviews by Roelof & Smart, this issue).

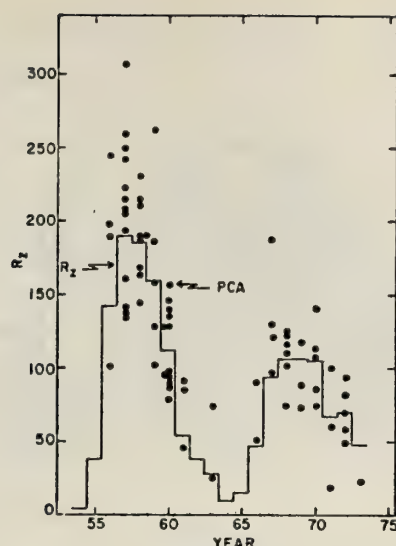


Figure 5.1 COMPARISON OF SUNSPOT NUMBER AT THE TIME OF EACH INDIVIDUAL PCA WITH THE ANNUAL MEAN  $R_z$  (SOLID LINE). (POMERANTZ & DUGGAL 1974).

### 5.1 Effects of SPEs on radio wave propagation

The effects of anomalous ionization over the polar regions on radio wave propagation has long been a subject of active research and international collaboration. Most attention has probably been given to the HF and VLF transpolar propagation, but propagation at LF and in recent years ELF has also been explored. For basic reference the review by Belrose (1968), and the books edited by Folkestad (1968b) on "Ionospheric Radio Communication" and Holtet (1974) on "ELF-VLF Radio Wave Propagation" are recommended. Polar problems in connection with HF communication are described by Lied (1967).

#### 5.1.1 Effects at VLF and LF

At VLF frequencies a number of studies have been carried out to establish the effects on propagation (Potemra et al 1967, 1969, Egeland & Naustvik 1967, Åbom et al 1969, Westerlund et al 1969, Oelberman 1970, Field & Greifinger 1972).

Transpolar VLF paths have proved to be a quite sensitive detector for SPEs and in Figure 5.2 is shown a histogram of the 73 cases of SPE activity detected by Westerlund (1974, private communication) between 1965-1973 by monitoring of circuits between the US and Kiruna, Sweden. The number of cases here is larger than in Figure 5.1, but the general positive correlation with mean annual sunspot number (solid line) is apparent. The classification into three classes of events is based upon proton flux data ( $>25$  MeV) recorded by the Vela satellites.

In their studies Westerlund et al (1969) concluded that transpolar VLF circuits were sensitive to the effects of proton fluxes ( $E_p > 25$  MeV) as low as  $\sim 0.5$  protons/cm<sup>2</sup> s ster. Figure 5.3 shows this result (in omnidirectional proton flux) for two types of propagation paths, one with high conductivity over the lower boundary, and one with a low conductivity such as across the Greenland icecap. Greater sensitivity for phase effects are observed on the latter type.



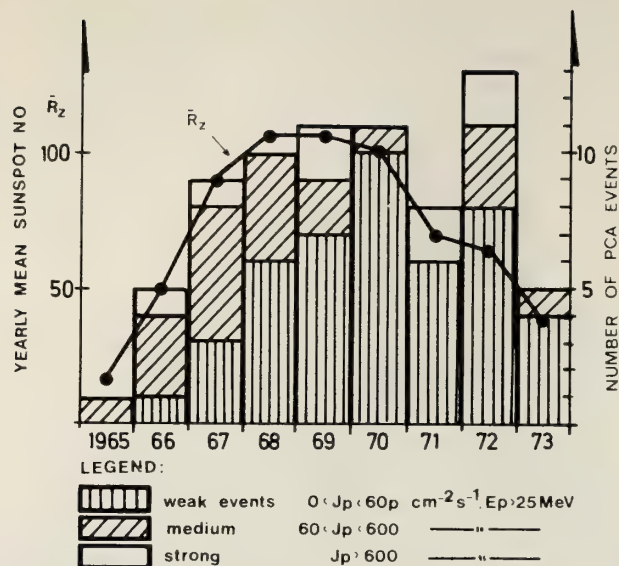


Figure 5.2 HISTOGRAM SHOWING NUMBER OF PCA EVENTS FOR THE YEARS 1965 TO 1973 (BASED ON DATA FROM WESTERLUND, PRIVATE COMMUNICATION, 1974). ALSO PLOTTED ARE THE YEARLY MEAN SUNSPOT NUMBERS.

Fitchtel and MacDonald (1967) found effects for  $0.5 \text{ protons/cm}^2 \text{ s ster}$  above 20 MeV and Oelbermann (1970) obtained VLF effects for fluxes of  $1 \text{ proton/cm}^2 \text{ s sr}$  above 20 MeV.

Oelbermann (1970) has reported the results of 5 years of monitoring phase and amplitude for polar and transpolar paths during 30 PCAs from 1965-69.

His main findings are:

- VLF phase (18.6 kHz) is severely advanced and amplitude is seriously attenuated during solar proton events. Phase changes are sometimes measured in cycles, attenuation is often greater than 30 dB.

- VLF field strength is a very sensitive detector of PCEs being sensitive to changes in proton flux on the order of  $1 \text{ proton/cm}^2 \text{-sec-ster}$  for protons of  $E > 20 \text{ MeV}$ .

- Nighttime transpolar VLF is 10-40 dB stronger than the daytime signal.

- Transpolar VLF signals are stronger in the winter than in the summer by 10-20 dB.

- An apparent increase in solar particle flux at sunspot maximum produces increased absorption of transpolar VLF which results in a reduction of signal strength of from 5 to 10 dB.

This latter finding is illustrated in Figure 5.4 which shows average daily amplitude variation (in absolute units) for periods of low and high solar activity as well as during SPEs.

For the NPG to Switzerland path the amplitude is:

- strongest at night, sunspot minimum
- weakest at day, sunspot maximum

Oelbermann (1970) has also detected heavy attenuation on paths which crosses the southern part of the polar cap where the auroral zone dips below the Arctic circle. Svennesson (1973) has made similar observations, cf. section 4.3.

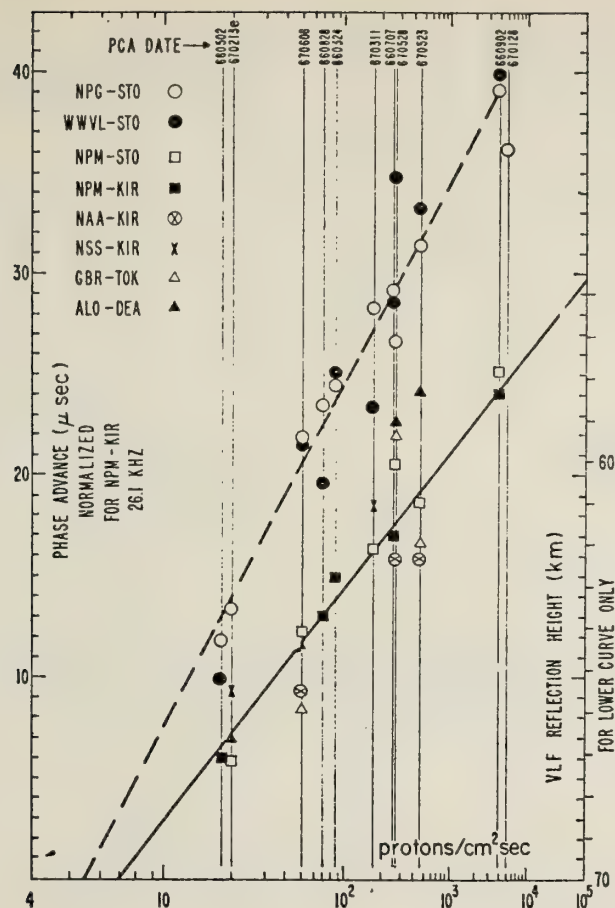


Figure 5.3 DAYTIME PHASE ADVANCE AS FUNCTION OF PROTON FLUX ( $E > 25$  MeV). PHASE ADVANCES ARE NORMALIZED TO 26.1 kHz (WAIT & SPIES 1964) AND TO A PCA-AFFECTED PATH PORTION OF 5300 KM. THE TWO LINEAR CURVES PERTAIN TO PATHS OVER GROUND OF HIGH  $\sigma_g$  AS REPRESENTED BY NPM-KIR (SOLID), AND PATHS OVER ABOUT 1000 KM OF GREENLAND ICE AS REPRESENTED BY NPC-STO (DASHED) (WESTERLUND ET AL 1969).

Westerlund et al (1969) has derived an expression for the phase deviations ( $\Delta\phi$ ) using the data present in Figure 5.3. Normalized to a frequency of 26.1 kHz and an affected path length of 5300 km (north of  $62.5^\circ$  corrected geomagnetic latitude, cf. Hakura (1965)), the expression for a path which crosses Greenland:

$$\Delta\phi = -12.9 + 24.5 \ln (\ln F/F_0) \quad F \geq 5.4$$

or for a sea water path:

$$\Delta\phi = -10.9 + 16.5 \ln (\ln F/F_0) \quad F \geq 7$$

where the  $F$  is the omnidirectional flux, and  $F_0 = 1$  proton/cm<sup>2</sup>s above 25 MeV. If their extrapolation to high solar flux levels is correct, an upper limit for the phase deviation for e.g. the path NPM, Seattle to Kiruna (26.1 kHz) would be ~25  $\mu$ s.

In a study of phase deviations at an Omega frequency of 10.2 kHz Swanson (1974) has derived a probability distribution function of the SPE in-

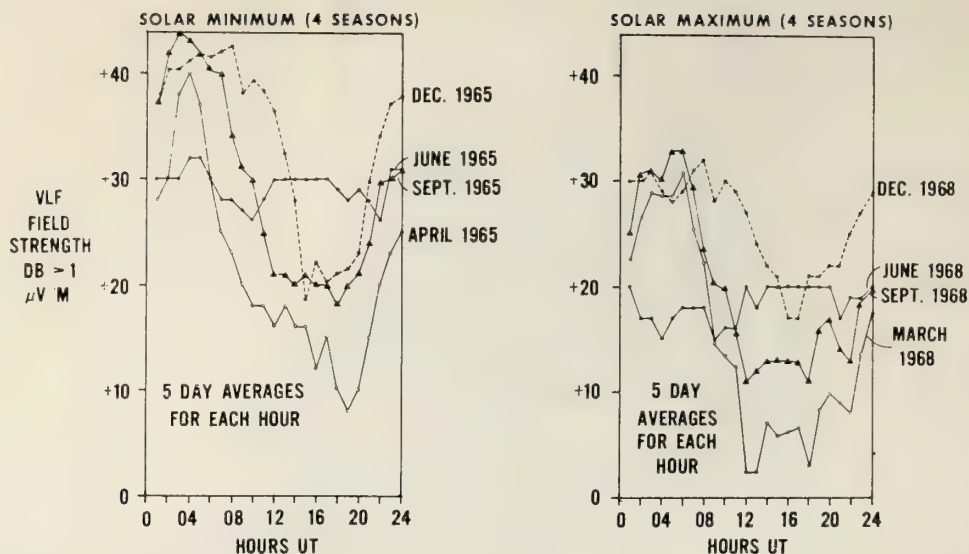


Figure 5.4 COMPARISON OF DIURNAL VLF DATA OF NPG-PAYERNE AT SOLAR MINIMUM AND SOLAR MAXIMUM FOR EACH SEASON (OELBERMAN 1970).

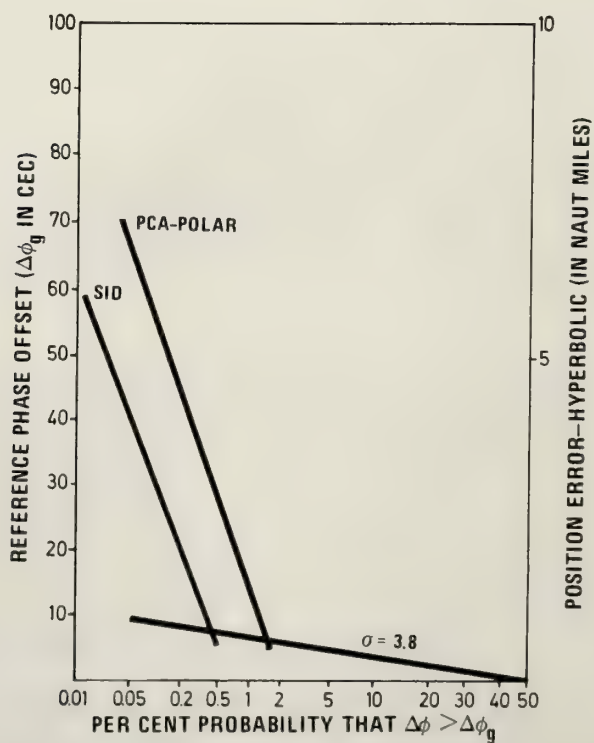


Figure 5.5 IMPORTANCE OF PCAs AND SIDs ON OMEGA PROPAGATION (10.2 kHz) AVERAGED OVER A SOLAR CYCLE (SWANSON 1974).



duced perturbations. Figure 5.5 shows results of data averaged over a solar cycle and is compared with similar data for SID events (cf. also Figure 2.3). Phase errors exceeding e.g. 30 centicycles ( $\sim 30 \mu\text{s}$ ) should on the average occur less than 0.5% of the time. Larsen (1977) found that for solar maximum activity the maximum mean phase deviation was 38 centicycles for the pair Omega Norway minus Omega Hawaii recorded at Svalbard ( $78^\circ\text{N}$ ). During the SPE of 2 November 1969 he deduced position errors of almost 10 km using Omega.

Efforts have been made to reduce the effects of such large positional errors. Rothmuller (1978) and Argo & Rothmuller (1978, this issue) has described an Omega correction model developed by Naval Ocean Systems Center which translates solar proton flux measured by satellites into phase error corrections. These corrections are in form similar to those now routinely used to account for variations during day, season, etc. The model has as parameters integrated proton flux  $>10 \text{ MeV}$ , time of day and path length through the polar cap (defined as north of  $63^\circ$  geomagnetic latitude). Figure 5.6 shows both corrected and uncorrected positional information for a solar proton event (Argo 1975). The model is seen to correct the phase deviations to give nominal 2 km positional accuracy.

In their studies Potemra et al (1967, 1969) have with good success compared experimental and calculated data on VLF phase deviations and HF absorption (see also Potemra 1972, Swider & Chidsey 1977). The riometer absorptions during a PCE event can be significant, exceeding the reliable range for the instrument ( $>20 \text{ dB}$ ). At 10 MHz absorption can amount to 100-200 dB, resulting

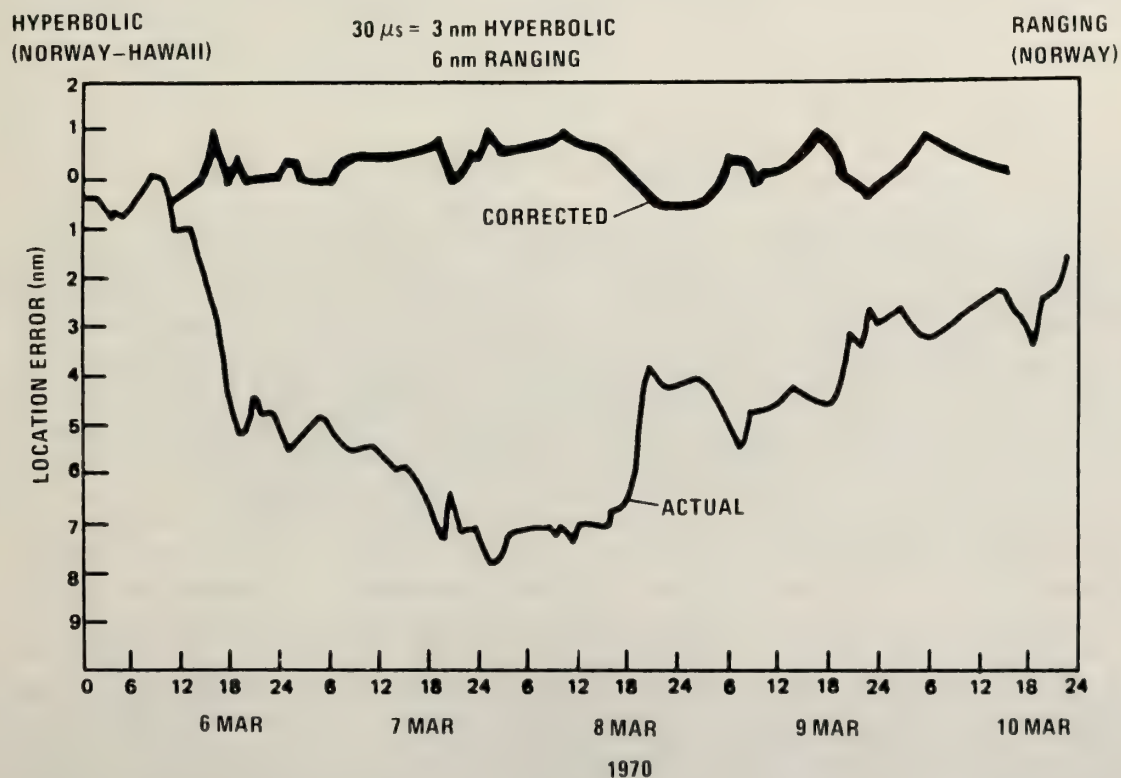


Figure 5.6 OMEGA LOCATION ERRORS (IN NAUTICAL MILES) DURING A SOLAR PROTON EVENT, ACTUAL AND CORRECTED VALUES (ARGO 1975).

in blackout period. Reid (1972) has deduced an empirical relationship between riometer absorption (A in dB) at 30 MHz and the  $2\pi$  omnidirectional proton flux (J) above 20 MeV:

$$J(>20 \text{ MeV}) = 60 A^2$$

Belrose (1968) has discussed at greath length LF data during disturbed conditions. Thrane & Larsen (1977) has described possible interference from Loran-C skywave upon the groundwave pulse during solar proton events. Due to a lowering of the reflection levels, the skywave may, at some distance from the transmitter, arrive before the navigation information is extracted from the pulse train.

#### 5.1.2 Effects at ELF

No actual signal measurements of stable ELF transmissions have been made during a solar proton event, but predictions of such effects can be made using available waveguide computer programs and data for previous SPE events.

The result of such a study (Larsen 1979, to be published) is that several dB signal attenuation is expected at 75 Hz, but actual amount depend critically upon the electron and ion density profiles. The details of the positive ion profile at low altitudes is of importance even down to 20-30 km. Attenuation rates are calculated to increase from ~0.9-1.7 dB/Mm to 2-3 dB/Mm for propagation at night over sea water ( $\sigma \sim 4 \text{ S/m}$ ) and ice ( $\sigma \sim 10^{-5} \text{ S/m}$ ), respectively. Simultaneously, the excitation factor also increases by several dB.

## 6. STRATOSPHERIC WARMINGS

During the local midwinter season sudden temperature changes in the stratosphere and mesosphere may occur over the polar and mid-latitude regions. These stratospheric warming (stratwarm) phenomena, first observed by Scherhag (1952) over Berlin, are synoptic scale events involving temporary changes in the global circulation system. The events start around the stratopause, which descends while the warming intensifies (Labitzke 1972a). During some events (major stratwarms) the polar vortex circulation pattern temporarily breaks down. During the major event in 1967/68 temperatures e.g. at the 37 km level were recorded to lie  $78^\circ\text{C}$  above the CIRA-1965 values.

The events extend into the mesosphere and Labitzke (1972b), referring to Russian rocket experiments at Thumba ( $9^\circ\text{N}$ ) and rocket granade experiments in North-America ( $\sim 71^\circ\text{N}$ ), has pointed out that temperature changes of opposite sign are found at strato- and mesosphere levels and likewise at high and low latitudes. The region of sign changes in the temperature is approximately at 50-55 km. From the mean vertical temperature profiles Labitzke (1972a) infers that the upper mesosphere is cold at high latitudes in early and late winter, and that the warm upper mesosphere (in northern hemisphere) observed in late January - early February is associated with the breakdown of the stratospheric polar vortex. Kriester (1966) has compiled characteristics for the stratwarms between 1952-1966; Finger et al (1974) for the period 1952-73.

The coupling between the ionosphere and the stratosphere was demonstrated by Bossolasco & Elena (1963) and Shapley & Beynon (1965) in their studies



of the correlation between ionospheric absorption and stratospheric warmings. The latter authors observed that ionospheric absorption increased a few days after a temperature increase in the stratosphere. Bowhill (1969) gives a review on these interactions with an extensive bibliography for literature published up to 1967. The more recent reviews by Labitzke (1975, 1977) also contain a good bibliographies.

## 6.1 Effects of stratwarms upon radio wave propagation

In addition to the mentioned absorption effects, additional effects on radio wave propagation have been detected by Belrose (1967) as lowering of VLF reflection levels and increase in MF vertical absorption. For the major warming in 1952, Belrose observed a lowering of 9-10 km of VLF reflection heights for a short propagation path (Rugby-Cambridge, UK). He concluded that the VLF and absorption A-figure of the non-deviative ionospheric absorption of the D- and E-regions were clear indications of changes in D-region ionization chemistry brought about by the stratwarm.

Belrose (1968) furthermore found for the January/February 1963 stratwarm that high HF absorption was correlated with greater heights and increased temperatures of the 10 mb level whereas the VLF reflections levels were lower than usual. However, high 10 mb temperatures are not always associated with high HF absorption (Belrose & Thomas 1968, Larsen 1971).

Variations in the 10 mb height levels have been correlated with long-distance VLF phase propagation data for daytime conditions (Cavalieri et al 1974). These authors invoke the upward propagation of energy from planetary scale waves in the stratosphere as providing the driving mechanism for the coupling to the ionosphere at 70 km giving rise to phase oscillations.

The D-region electron densities increased during the 1952 event (Belrose 1967). For an event in 1967/68 Rowe et al (1969) measured increases in electron densities at 80 and 90 km by factors of 10 and 8, respectively, while no significant change was observed at 70 km. During a minor stratwarm in 1969 Larsen (1971) deduced an increase in VLF reflection height of 3 km for a short VLF path at high latitudes, indicating a depletion in electron density around 60 km.

At present, one can conclude that stratospheric warmings do give rise to effects in radio wave propagation characteristics at middle and high latitudes. The change imposed on the free electron density are shown to be height dependent, and probably also have significant spatial variations, as well as differences are expected from event to event.

## 7. CONCLUDING REMARKS

The D-region effects on radio wave propagation constitute a many-faceted problem area. Within the limited space available only the major effects could be treated, and the selection of examples discussed are probably biased by the author's interest. An apology is extended to authors of papers which are relevant for an extensive discussion of these phenomena, but which here are omitted or only briefly mentioned. A conscientious reader of the references given and their bibliographies will, however, obtain a fairly complete overview of this interesting area of science.



## REFERENCES

- Agy, V. (1970): Ed. of "Ionospheric Forecasting", AGARD Conf. Proc. 46, NATO-AGARD.
- Aikin, A.C. (1974): Energetic particle influence in the mid-latitude D-region in winter, EOS Trans. AGU, 55:1160.
- Apostolov, E. (1975): Seasonal dependence of SFA effects at 164 kHz resulting from structural changes in the D-region, B'ulgarska Akademiia na Naukite, Geofizichni Institut, 20:85. (In Bulgarian.)
- Appleton, E.V., and W.R. Piggott (1954): Ionospheric absorption measurements during a solar cycle. J. Atmosph. Terr. Phys., 5:141.
- Argo, P.E. (1975): Modeling Omega PCA phase advances, Techn. report TR 1950, Naval Electronics Laboratory Center, San Diego, CA., USA.
- Argo, P.E., and I.J. Rothmuller (1978): Prophet: An application of propagation forecasting principles, ISTP preprint No 13.
- Bailey, D.K. (1968): Some quantitative aspects of electron precipitation in and near the auroral zone, Rev. Geophys., 6:289.
- Bailey, D.K., and M.A. Pomerantz (1965): Relativistic electron precipitation into the mesosphere at subauroral latitudes, J. Geophys. Res., 70:5823.
- Bain, W.C., and E. Hammond (1975): Ionospheric solar flare effect observations, J. Atmosph. Terr. Phys., 37:573.
- Bain, W.C., and B.R. May (1967): D-region electron-density distributions from propagation data, Proc. IEEE, 114:1593.
- Bannister, P.R. (1975): Variations in extremely low frequency propagation parameters, J. Atmosph. Terr. Phys., 37:1203.
- Barclay, L.W. (1978): Statistical modelling of HF links, AGARD Conf on "Operational modelling of the aerospace propagation environment", AGARD-CPP-238, NATO-AGARD.
- Barr, R. (1977): The effect of sporadic-E on the nocturnal propagation of ELF radio waves, J. Atmosph. Terr. Phys., 39:1379.
- Bates, H.F., and P.R. Albee (1965): General VLF phase variations observed at College, Alaska, J. Geophys. Res., 70:2187.
- Belikovich, V.V., E.A. Benediktov, and M.A. Itkina (1971): Results of determinations of the electron density profile in the ionospheric D-region from the frequency dependence of radio wave absorption, Geomagnetism and Aeronomy, 11:222.

- Belikovich, V.V., E.A. Benediktov, L.V. Grishkevich, and V.A. Ivanov (1975): Results of electron concentration measurements in the ionospheric D-region during ionospheric disturbances, Radiofizika, 18:1094 (In Russian).
- Belikovich, V.V., E.A. Benedictov, and M.A. Itkina (1977): The function of electron loss in the D-region of the ionosphere and the dependence of anomalous radio wave absorption on the angle of the solar zenith during sudden ionospheric disturbances, Geomagnetism de Aeronomia, 17:427 (In Russian).
- Belrose, J.S. (1967): The "Berlin" warming, Nature, 214:660.
- Belrose, J.S. (1968): Low and very low frequency radio wave propagation, In: AGARD lecture series on "Radio Wave Propagation", AGARD-LS-XXIX, NATO-AGARD.
- Belrose, J.S. (1969): The winter variability of electron density in the middle latitude D-region. In "Progress in Radio Science 1966-1969", Proc. URSI General Assembly, Vol. I, Ottawa, Canada, 65.
- Belrose, J.S., and L. Thomas (1968): Ionization changes in the middle latitude D-region associated with geomagnetic storms, J. Atmos. Terr. Phys., 30:1397.
- Berkey, F.T., V.M. Driatskiy, K. Henriksen, D.H. Jelly, T.I. Shuchuka, A. Theander, and J. Yliniemi (1971): Temporal development of the geographical distribution of auroral absorption for 30 substorm events in each of IQSY (1964-65) and IASY (1969), Report UAG-16, World Data Center A.
- Berkey, F.T., V.M. Driatskiy, K. Henriksen, B. Hultqvist, D.H. Jelly, T.I. Shuhuka, A. Theander, and J Ylimiemi (1974): A synoptic investigation of particle precipitation dynamics for 60 substorms in IQSY (1964-65) and IASY (1969), Planet. Space Sci., 22:255.
- Beynon, W.J.G., and E.R. Williams (1976): Rocket measurements of D-region electron density profiles, J. Atmosph. Terr. Phys., 38: 1319.
- Beynon, W.J.G., E.R. Williams, F. Arnold, D. Krankowsky, W.C. Bain, and P.H.G. Dickinson (1976): D-region rocket measurements in winter anomaly absorption conditions, Nature, 261:118.
- Blake, J.B., and A.L. Vampola (1974): A study of magnetospheric boundaries by the simultaneous observations of solar and magnetospheric particles. In: Correlated interplanetary and magnetospheric observations, edited by D.E. Page, D. Reidel Publ. Co., Dordrecht-Holland, 173.
- Bleiweiss, M.P., V.E. Hildebrand, and J.R. Hill (1973): A D-region model which accounts for quiet and disturbed VLF propagation phenomena, Technical Report 1868, Naval Electronics Lab. Center, San Diego, Ca., USA.

- Booker, H.G. (1977): The relation between ionospheric profiles and ELF propagation in the earth-ionosphere transmission line, Final report 1 Apr 75 - 30 Nov 76, Univ. of Ca., San Diego, CA., USA.
- Booker, H.G., and F. Lefeuvre (1977): The relation between ionospheric profiles and ELF propagation in the Earth-ionosphere transmission line, J. Atmosph. Terr. Phys., 39:1277.
- Bossolasco, M., and A. Elena (1963): Absorption de la couche D et température de la mésosphère, Compt. Rend., 256:4491.
- Bowhill, S.A. (1969): Interactions between the stratosphere and the ionosphere. In Annals of the IQSY, 5, MIT Press, Cambridge, USA, 83.
- Bracewell, R.N., K.G. Budden, J.A. Ratcliffe, T.W. Straker, and K. Weekes (1951): The ionospheric propagation of low- and very-low-frequency radio waves over distances less than 1000 km, 3, Proc. Inst. Elec. Eng., 98:221.
- Burgess, B., and T.B. Jones (1967): Solar flare effects and VLF radio wave observations of the lower ionosphere, Radio Science, 2:619.
- Burgess, B., and D. Walker (1970): Effects in Omega from propagation variations, J. Inst. of Navigation, 23:49.
- Cavalieri, D.J., R.J. Deland, T.A. Potemra, and R.F. Gavin (1974): The correlation of VLF propagation variations with atmospheric planetary-scale waves, J. Atmosph. Terr. Phys., 36:561.
- Chakravarty, S.C., and K.R. Ramanathan (1972): A comparison of field-strengths of 164 kHz radio waves transmitted from Tashkent and received at Ahmedabad with flare-time solar X-ray emissions measured in satellites, Proc. Indian Academy of Sciences, Section A, 75: 249.
- Chilton, C.J., J.P. Conner, and F.K. Steele (1965): A comparison between solar X-ray emission and VLF sudden phase anomalies, Proc. IEEE, 53:2018.
- Cipriano, J.P., L.C. Hale, and J.D. Mitchell (1974): Relations among low ionosphere parameters and A3 radio wave absorption, J. Geophys. Res., 79:2260.
- Dieminger, W. (1952): Über de Ursache der exgessiven Absorption in der Ionosphäre an Wintertagen, J. Atmosph. Terr. Phys., 23:340.
- Donnelly, R.F.(1968): ESSA Technical report ERL92-SDL6, ESSA, Boulder, Co, USA.
- Donnelly, R.F. (1976): Solar flare X-ray and EUV emission; A terrestrial viewpoint. In Physics of solar planetary environments, Proc. of the International Symposium on Solar-Terrestrial Physics, edited by D.J. Williams, American Geophysical Union, 178.



- Driatskiy, V.M., T.M. Krupitskaia, and A.V. Shirochov (1971): Determination of absorption in the lower ionosphere from data on the solar cosmic ray flux density and spectrum, Geomagnetism and Aeronomy, 11:608 (Translation).
- Davies, K. (1978a): Ionospheric prediction and extrapolation. In AGARD Proc. on "Operational modelling of the aerospace propagation environment", AGARD Conf. Preprint, CPP-238, NATO-AGARD.
- Davies, K. (1978b): Forecasting and prediction of ionospheric parameters. In AGARD lecture series on "Recent advances in radio and optical propagation for modern communications, navigation and detection systems", AGARD-LS-93, NATO-AGARD.
- Davis, J.R. (1976): Localized nighttime D-region disturbances and ELF propagation, J. Atmosph. Terr. Phys., 38:1309.
- Desphande, D.S., and A.P. Mitra (1972): Ionospheric effects of solar flares-III. The quantitative relationship of flare X-rays to SID's, J. Atmosph. Terr. Phys., 34:243.
- Egeland, A., and E. Naustvik (1967): Influence of high-latitude disturbances on VLF propagation, Radio Science, 2 (new series), 659.
- Egeland, A, T.R. Larsen, and E. Naustvik (1969): The position and height of auroral absorption deduced from VLF measurements, J. Atmosph. Terr. Phys., 31:187.
- Elling, W., and H. Schwentek (1977): Über Perioden in der Absorption von Radiowellen in der Ionosphäre, Kleinheubachter Berichte, 21:257 (In German).
- Evans, L.C., and E.C. Stone (1969): Access of solar protons into the polar cap: A persistent north-south asymmetry, J. Geophys. Res., 74:5127.
- Field, E.C., C. Greifinger, and K. Schwartz (1972): Transpolar propagation of long radio waves, J. Geophys. Res., 77:1264.
- Fichtel, C.E., and F.B. McDonald (1967): Energetic particles from the sun, A. Rev. Astron. Astrophys., 5:351.
- Folkestad, K. (1968a): Results from stepped frequency oblique soundings at high latitudes. In Ionospheric Radio Communications, edited by K Folkestad, Plenum Press, New York, 279.
- Folkestad, K. (1968b), ed. In Ionospheric Radio Communication, Plenum Press, New York.
- Ganguly, S. (1972): Some results of absorption measurements at a low latitude station, J. Atmosph. Terr. Phys., 34:2009.
- Ganguly, S. (1975): Particulate effects on the low latitude D-region, J. Geophys. Res., 80:2883.

- Gough, M.P., and H.L. Collin (1973): Energetic electron precipitation as a source of ionization in the night-time D-region over the mid-latitude rocket range, South Uist., J. Atmosph. Terr. Phys., 35:835.
- Gregory, J.B., and A.H. Manson (1970): Seasonal variations of electron densities below 100 km at mid-latitudes, J. Atmosph. Terr. Phys., 32:837.
- Hakura, Y. (1965): Initial phase of the PCA events and compositions of the solar cosmic radiations, Japan. J. Radio Res. Lab., 12:231.
- Hakura, Y. (1970): Mapping of the polar cap ionosphere by solar and magnetosphere particles, J. Franklin Inst., 290:263.
- Hartz, T.R., and N.M. Brice (1967): The general pattern of auroral particle precipitation, Planet. Space Sci., 15:301.
- Hartz, T.R., L.E. Montbriand, and E.L. Vogan (1963): A study of auroral absorption at 30 Mc/s, Canad. J. Phys., 41:581.
- Holtet, J. (1974): Ed., ELF Radio Wave Propagation, D. Reidel Publ. Co., Dordrecht-Holland/Boston, USA.
- Hargreaves, J.K. (1973): Mid-latitude winter anomalies in radio absorption and stratospheric temperature distribution, J. Atmosph. Terr. Phys., 35:291.
- Hargreaves, J.K., and F.C. Cowley (1967): Studies of auroral radio absorption events at three magnetic latitudes, Parts I and II, Planet. Space Sci., 15:71.
- Hidalgo, M.A. (1977): The winter anomaly in ionospheric absorption and the D-region ion chemistry, Planet. Space Sci., 25:1135.
- Imhof, W.L., T.R. Larsen, J.B. Reagan, and E.E. Gaines (1976): Analysis of satellite data on precipitating particles in coordination with ELF propagation anomalies, LMSC-D502063, Lockheed Palo Alto Research Laboratory, Palo Alto, Ca., USA.
- Imhof, W.L., T.R. Larsen, J.B. Reagan, and E.E. Gaines (1977): Analysis of satellite data on precipitating particles in coordination with ELF propagation anomalies, LMSC-D560323, Lockheed Palo Alto Research Laboratory, Palo Alto, Ca., USA.
- Imhof, W.L., J.B. Reagan, E.E. Gaines, T.R. Larsen, J.R. Davis, and W. Moler (1978): Coordinated measurements of ELF transmission anomalies and the precipitation of energetic particles into the ionosphere, Radio Science, 13:717.
- Jean, A.G., and D.B. Large (1967): ITSA Technical Memo No. 90, Institute for Telecommunication Sciences and Aeronomy, Boulder, Co., USA.

- Jones, J., and J.G. Collins (1973): The effect of D-region absorption on the forward-scatter radio meteor amplitude distribution index, J. Atmosph. Terr. Phys., 35:2277.
- Jones, T.B. (1967): Ionospheric effects of solar X-ray enhancements. In AGARD Proc. on "Propagation Factors in Space Communication", AGARD-CP-3, NATO-AGARD.
- Jones, T.B. (1978a): The propagation of low and very low frequency radio waves. In AGARD lecture series on "Recent Advances in Radio and Optical Propagation for Modern Communications, Navigation and Detection Systems", AGARD-LS-93, NATO-AGARD.
- Jones, T.B. (1978b): High frequency radio wave propagation in the ionosphere. In AGARD lecture series on "Recent Advances in Radio and Optical Propagation for Modern Communications, Navigation and Detection Systems", AGARD-LS-93, NATO-AGARD.
- Kalinovskaia, G.P., and L.P. Anischenko (1973): Anomalous winter absorption of radio waves in the mid-latitude ionosphere, Geomagnetism and Aeronomy, 13:315 (Translation).
- Kaufmann, P., and I. Rizzo Piazza (1978): On the relation of SPA measured at VLF to solar microwave burst energies - sudden phase anomalies, Solar Phys., 57:479.
- Komrakov, G.P., I.A. Skrebkova, and A.V. Tolmacheva (1972): Absorption of radio waves in the ionosphere, Ionosfernye Issledovania, 20:55 (In Russian).
- Kriester, B. (1966): Stratospheric warmings, Presented at NATO Advanced Study Institute on "Movements and turbulence of the atmosphere between 30 and 120 km", Lindau 19 Sept - 1 Oct 1966.
- Labitzke, K. (1972a): Temperature changes in the mesosphere and stratosphere connected with circulation changes in winter, J. Atmosph. Sciences, 29:756.
- Labitzke, K. (1972b): The interaction between stratosphere and mesosphere in winter, J. Atmosph. Sciences, 29:1395.
- Labitzke, K. (1975): Review of investigations in the field: Meteorology of the stratosphere and mesosphere. In Space Research XV, Akademie-Verlag, Berlin, 81-108.
- Labitzke, K. (1977): Stratospheric - mesospheric midwinter warmings. In Dynamical and Chemical coupling between the neutral and ionized Atmosphere, Ed.: B. Grandal and J.A. Holtet, D. Reidel Publishing Company, Dordrecht-Holland/Boston-USA, 17-34.
- Laštovička, J. (1975): On the role of atmospheric pressure and density variations in ionospheric absorption of radio waves, J. Atmosph. Terr. Phys., 37:377.



- Laštovička, J. (1976): The dependence of the ionospheric absorption at 2775 kHz on the intensity of ionizing radiation - Ionospheric implications, Pure and Appl. Geophys., 114:321.
- Laštovička, J., and J. Smilauer (1973): Satellite measurements of solar X-ray flux and ground observations of sudden ionospheric disturbances, Geofysikalni Sbornik, 21:305.
- Lanzerotti, L.J. (1972): Solar energetic particles and the configuration of the magnetosphere, Rev. Geophys. Space Res., 10:379.
- Lanzerotti, L.J., and C.G. MacLennan (1972): Relative importance of solar electrons, protons and alphas in the November 1969 PCA event, In Proc. COSPAR Symposium on Solar Particle Event of November 1969, edited by J.C. Ulwick, 85.
- Larsen, T.R. (1971): Short path VLF phase and amplitude measurements during a stratospheric warming in February 1969, J. Atmosph. Terr. Phys., 33:1251.
- Larsen, T.R. (1973): Disturbances in the high latitude, lower ionosphere, NDRE Report No. 62, Norwegian Defence Research Establishment, Kjeller, Norway.
- Larsen, T.R. (1974): Irregular variations in the high latitude ionosphere and their effects on propagation. In ELF-VLF Radio Wave Propagation, edited by J. Holtet, D. Reidel Publishing Co., Nordrecht-Holland/Boston-USA, 171.
- Larsen, T.R. (1977): Omega accuracy in polar regions during ionospheric disturbances. In AGARD Conf. Proc. on "Propagation Limitations of Navigation and Positioning Systems", AGARD-CP-209, NATO-AGARD.
- Larsen, T.R., T.A. Potemra, W.L. Imhof, and J.B. Reagan (1977): Energetic electron precipitation and VLF phase disturbances at middle latitudes following the magnetic storm of December 16, 1971, J. Geophys. Res., 82:1519.
- Larsen, T.R., J.B. Reagan, W.L. Imhof, L.E. Montbriand, and J.S. Belrose (1976): A coordinated study of energetic electron concentrations over Ottawa during disturbed conditions, J. Geophys. Res., 81:2000.
- Larsen, T.R., and G.R. Thomas (1974): Energy spectra measured during a relativistic electron precipitation event on 2 February 1969, J. Atmosph. Terr. Phys., 36:1613.
- Larsen, T.R., and E.V. Thrane (1977): Ionospheric effects on Loran-C in polar regions, In AGARD Conf. Proc. on "Propagation Limitations of Navigation and Positioning Systems", AGARD-CP-209, NATO-AGARD.
- Lauter, E.A. (1961): Excessive ionization in the lower ionosphere and instabilities in the outer radiation belt, Naturwissenschaften, 48:473.

- Lauter, E.A., and R. Knuth (1967): Precipitation of high energy particles into the upper atmosphere at medium latitude after magnetic storms, J. Atmosph. Terr. Phys., 29:411
- Lavergnat, J., J.J. Berthelmer, M. Sylvain, and J. Vassal (1976): Discussion critique de l'application des mesures-d'absorption par riometre, a l'etude de l'ionosphere, Annals de Geophysique, 32:391.
- Leonov, V.I. (1975): Characteristics of VLF propagation on auroral paths and the spatial-temporal properties of auroral particle fluxes. In Problemy Difraktsii i Rasprostraneniia, 14:109 (In Russian).
- Leroy, B. (1978): Relations between solar X-ray bursts, sudden enhancements of atmospherics and regions of solar activity, ISTP Preprint No. 56.
- Lied, F. (1967): Ed. High frequency radio communications with emphasis on polar problems, AGARDograph 104, Technivision, Maidenhead, England.
- Matthews, D.L., and D.J. Simons (1973): Observation of relativistic electron precipitation at L=6, J. Geophys. Res., 78:7539.
- Manson, A.H., and M.W.J. Merry (1970): Particle influx and the "winter anomaly" in the mid-latitude (L=2,5-3,5) lower ionosphere, J. Atmosph. Terr. Phys., 32:1169.
- Montbriand, L.E., and J.S. Belrose (1972): Effective electron loss rates in the lower D-region during the decay of solar X-ray flare events, Radio Sci., 7:133.
- Montbriand, L.E., and J.S. Belrose (1976): Changes in electron precipitation inferred from spectra deduced from D-region electron densities during a postmagnetic storm, J. Geophys. Res., 81:2213.
- Mitra, A.P. (1974): Ionospheric effects of solar flares, D. Reidel Publishing Company, Dordrecht-Holland/Boston-USA.
- Oelbermann, Jr, E.J. (1970): Solar particle effects on polar cap VLF propagation, J. Franklin Inst., 290:281.
- Offerman, D. (1977): Some results from the European Winter Anomaly Campaign 1975-76. In Dynamical and chemical coupling of neutral and ionized atmosphere, edited by B. Grandal and J.A. Holtet, D. Reidel Publ. Co., Dordrecht-Holland/Boston-USA, 235.
- Oyinloye, J.O. (1975): Radio wave absorption in the equatorial ionosphere, J. Atmosph. Terr. Phys., 37:1.
- Nestorov, G.T. (1977): Signal variations of LF radio-waves over the sunspot cycle, J. Atmosph. Terr. Phys., 39:741.
- Paulikas, G.A. (1975): Precipitation of particles at low and middle latitudes, Rev. Geophys. Space Phys., 13:709.



- Perry, M.E. (1977): A study of sudden ionospheric disturbances and their effects on VLF position fixing accuracy. In AGARD Conf. Proc. "Propagation limitations of navigation and positioning systems", AGARD-CP-209, NATO-AGARD.
- Pomerantz, M.A., and S.P. Duggal (1974): The sun and cosmic rays, Rev. Geophys. Space Phys., 12:343.
- Potemra, T.A. (1972): The empirical connection of riometer absorption to solar protons during PCA events, Radio Sci., 7:571.
- Potemra, T.A. (1973): Precipitating energetic electrons in the mid-latitude lower ionosphere. In Physics and Chemistry of the Upper Atmosphere, edited by B.M. McCormac, D. Reidel, Dordrecht, Netherlands, 67.
- Potemra, T.A., A.J. Zmuda, C.R. Haave, and B.W. Shaw (1967): VLF phase perturbations produced by solar protons in the event of February 5, 1965, J. Geophys. Res., 72:6077.
- Potemra, T.A., A.J. Zmuda, C.R. Haave, and B.W. Shaw (1969): VLF phase disturbances, HF absorption and solar protons in the events of August 28 and September 2, 1966, J. Geophys. Res. (Space Physics), 74:6444.
- Potemra, T.A., and A.J. Zmuda (1972): Solar electrons and alpha particles during Polar-Cap absorption events, J. Geophys. Res., 77:6916.
- Potemra, T.A., and T.J. Rosenberg (1973): VLF propagation disturbances and electron precipitation at mid-latitudes, J. Geophys. Res., 78:1572.
- Ranta, H., and A. Ranta (1977): Daily variation of absorption in the D-region using data at high latitudes, J. Atmosph. Terr. Phys., 39:309.
- Rangasawamy, S. (1974): Electron density profiles deduced from absorption measurements, Indian J. Radio Space Phys., 3:104.
- Rawer, K. (1975): The historical development of forecasting methods for ionospheric propagation of HF waves, Radio Science, 10:669.
- Reagan, J.B. (1977): Ionization process. In Dynamical and chemical coupling between the neutral and ionized atmosphere, Editors: B. Grandal and J.A. Holtet, D. Reidel Publishing Company, Dordrecht-Holland&Boston-USA, 145-160.
- Reagan, J.B. (1975): A study of the D-region ionosphere during the intense solar particle events of August 1972, LMSC-D454290, Lockheed Palo Alto Res. Lab., Palo Alto, Ca., USA.
- Reagan, J.B., W.L. Imhof, and E.E. Gaines (1972): Satellite measurements of energetic solar protons, alpha particles, electrons and auroral particles during the 2 November 1969 PCA event. In Proc. COSPAR Symposium on Solar Particle Event of November 1969, edited by J.C. Ulwick, AFCRL-72-0474, 141.



- Reder, F.H., and S. Westerlund (1970): VLF signal phase instabilities produced by propagation medium: Experimental results. In Phase and frequency instabilities in electromagnetic wave propagation, edited by K. Davies, Technivision Services, Slough, England, 103.
- Reder, F.H., C.J. Åbom, and G.M.R. Winkler (1964): Precise phase and amplitude measurements on VLF signals propagated through the Arctic zone, Radio Sci., J. Res. NBS 68D: 275.
- Reid, G.C. (1972): A review of ionospheric radio propagation effects associated with solar proton events. In Proc. COSPAR Symposium on Solar Particle Event of November 1969, edited by J.C. Ulwick, AFCRL-72-0474, 201.
- Reilly, A.E. (1977): Analysis of sweep frequency oblique polar region high frequency radio propagation measurements, Technical report AFGL-TR-77-0102, ERP No. 596, Air Force Geophysical Laboratory, Hanscom Air Force Base, Mass, USA.
- Rose, R.B., D.G. Morfitt, and M.P. Bleiweiss (1971): System performance degradation due to varying solar emission activity: SOLRAD application study, Tash II, Techn. report TR 1774, Naval Electronics Laboratory Center, San Diego, CA., USA.
- Rose, G., and H.U. Widdel (1977): D-region radio wave propagation experiments, their significance and results during the Western European Winter Anomaly Campaign 1975/76, J. of Geophysics, 44:15.
- Rosenberg, T.J., L.J. Lanzerotti, D.K. Bailey, and J.D. Pierson (1972): Energy spectra in relativistic electron precipitation events, J. Atmosph. Terr. Phys., 34:1977.
- Rothmuller, I.J. (1978): Real-time propagation assessment. In AGARD Proc. on "Operational Modelling of the Aerospace Propagation Environment", AGARD Conf. preprint AGARD-CPP-238, NATO-AGARD.
- Rowe, J.N., A.J. Ferraro, H.S. Lee, and A.P. Mitra (1969): Changes in electron density and collision frequency at University Park, Pennsylvania during the stratospheric warmings of 1967/68, J. Atmosph. Terr. Phys., 31:1077.
- Rowe, J.N., A.J. Ferraro, H.S. Lee, R.W. Kreplin, and A.P. Mitra (1970): Observations of electron density during a solar flare, J. Atmosph. Terr. Phys., 32:1609.
- Samra, S.B.S.S., and M.C. Shamra (1972): Cosmic noise absorption studies at Dehli over a solar cycle, Planet. Space Sci., 20:253.
- Samuel, J.C., and P.A. Bradley (1975): A new form of representation of the diurnal and solar-cycle variations of ionospheric absorption, J. Atmosph. Terr. Phys., 37:131.
- Sao, K., M. Yamashita, S. Tanahashi, H. Jindoh, and K. Ohta (1970): Sudden enhancements (SEA) and decreases (SDA) of atmospherics, J. Atmosph. Terr. Phys., 32:1567.

- Scherhag, R. (1952): Die explosionsartigen Stratosphärenwärmungen des Spät-winters 1951/52, Ber. Deut. Wetterdienst (US Zone), 38:51.
- Schwentek, B. (1971): The sunspot cycle 1958/70 in ionospheric absorption and stratospheric temperature, J. Atmosph. Terr. Phys., 33:1839.
- Seera, G.S., R.C. Dubey, and C.S.G.K. Setty (1973): Variations of D-region absorption at Dehli, Indian J. Radio and Space Phys., 2:119.
- Sengupta, P.R. (1974): A study of SFA on 164 kHz for Tashkent-Dehli propa-gation - sudden field anomaly due to solar X-ray flare. In International Symposium on Solar-Terrestrial Physics, Sao Paulo, Brazil, June 17-22, 1974, Proceedings Volume 3.
- Shain, C.A., and A.P. Mitra (1954): Effects of solar flares on the absorption of 18.3 Mc/s cosmic noise, J. Atmosph. Terr. Phys., 5:316.
- Shapley, A.H., and W.J.G. Beynon (1965): Winter anomaly in ionospheric ab-sorption and stratospheric warmings, Nature, 206:1242.
- Singer, W., and J. Bremer (1974): Mid-latitude D-region electron density pro-files at different solar activities adapted to LF, MF and HF propagation data, Indian J. Radio Space Phys., 3:119.
- Spjeldvik, W.N., and R.M. Thorne (1975): The cause of storm after effects in the middle latitude D-regions, J. Atmosph. Terr. Phys., 31:777.
- Svennesson, J. (1973): Effects on VLF propagation during ionospheric sub-storms, J. Atmosph. Terr. Phys., 35:761.
- Švestka, Z., and P. Simon (editors) (1975): Catalog of solar particle events 1955-1969, D. Reidel Publishing Company, Dordrecht-Holland/Boston-USA.
- Swanson, E.R. (1971): Omega, Navigation, 18:168.
- Swanson, E.R. (1974): Blunders caused by Omega propagation: SPA's and PCA's. Proc. 2nd Omega Symposium, 5-7 November 1974, Wash. D.C., 202.
- Swanson, E.R. (1977): Propagation effects on Omega. In AGARD Conf. Proc. "Propagation limitations of navigation and positioning systems", AGARD-CP-209, NATO-AGARD.
- Swanson, E.R., and C.P. Kugel (1973): A synoptic study of sudden phase anomalies (SPAs) effecting VLF navigation and timing, Proc. of the fifth annual NASA and Department of Defence, Precise Time and Time Interval (PTTI) Planning Meeting, NASA Publication X-814-74-225, 443.
- Swider, W., and I.L. Chidsey, Jr. (1977): Calculated and measured HF/VHF absorption for the 2-5 November 1969 solar proton event, Tech. report 76-0053, Air Force Geophysical Lab., Hanscom AF Base, Mass., USA.
- Theander, A. (1972): Studies of auroral absorption substorms, Report No. 723, Kiruna Geophysical Observatory, Kiruna, Sweden.



- Thomas, L. (1962): The winter anomaly in ionospheric absorption. In Radio Wave Absorption in the Ionosphere, edited by N.C. Gerson, Pergamon, New York, 301.
- Thomas, L. (1968): The electron density distribution in the D- and E-region during days of anomalous radio wave absorption in winter, J. Atmosph. Terr. Phys., 30:1211.
- Thorne, R.M., and T.R. Larsen (1976): An investigation of relativistic electron precipitation events and their association with magnetospheric sub-storm activity, J. Geophys. Res., 81:5501.
- Thrane, E.V. (1978): Geophysical disturbance effects and their predictability. On AGARD Conf. Proc. on "Operational Modelling of the Aerospace Propagation environment", AGARD Conf. Preprint, AGARD-CPP-238, NATO-AGARD.
- Thrane, E.V., W. Bangert, D. Berau, M. Friedrich, B. Grandal, O. Hagen, A. Loidl, K. Spenner, H. Schwentek, K.M. Torkar, and F. Ugletveit (1979): Ion production and effective loss rate in the ionosphere and lower thermosphere during the Western Europe Anomaly Campaign 1975-76, J. Atmosph. Terr. Phys., in press.
- Torkar, K.M. (1977): Die untere Ionosphäre und ihre Bedeutung um Energiehaushalt der hohen Atmosphäre unter besonderer Berücksichtigung der Messung von Elektronendichten und Stosszahlen mit Raketen, INW 7709, Inst. für Nachrichtentechnik und Wellenausbreitung, Technische Universität Graz, Austria (In German).
- Triska, P., and J. Laštovička (1972): Sudden decrease of atmospherics at 5 kHz, J. Atmosph. Terr. Phys., 34:1065.
- Ulwick, J.C. (1969): Ed., Proc. COSPAR Symposium on Solar Particle Event of November 1969, AFCRL-72-0474, Air Force Cambridge Res. Lab., Bedford, USA.
- Vampola, A.L. (1969): Energetic electrons at latitudes above the outer zone cutoff, J. Geophys. Res., 74:1254.
- Van Allen, J.A., J.F. Fennell, and N.F. Ness (1971): Asymmetric access of energetic solar protons to the earth's north and south polar caps, J. Geophys. Res., 76:4262.
- Veastad, J.R. (1968): Survey of the user problems and existing communication facilities in the Norwegian Arctic. In Ionospheric Radio Communications, edited by K. Folkestad, Plenum Press, New York, 422.
- Wait, J.R., and K.P. Spies (1964): Characteristics of the earth ionosphere waveguide for VLF radio waves, Techn. note No. 300, National Bureau of Standards, Boulder, Co., USA.
- Wakai, N., S. Fujii, and C. Ouchi (1973): SID effects as observed in intensities of LF radio waves, Japan. J. Radio Res. Lab., 20:67.



- Westerlund, S., F.H. Reder, and C. Åbom (1969): Effects of polar cap absorption events on VLF transmissions, J. Atmosph. Terr. Phys., 17:1329.
- Westerlund, S., and F.H. Reder (1973): VLF propagation at auroral latitudes, J. Atmosph. Terr. Phys., 35:1453.
- Yamashita, M. (1969): The conductivity of lower ionosphere deduced from sudden enhancements of strength (SES) of VLF transmissions, J. Atmosph. Terr. Phys., 31:1049.
- Zelenkova, L.V. (1976): Distribution of electron concentration in the D-region, Geomagnitnye Issledovaniia, 18:101 (In Russian).
- Zelenkova, V.L., O.I. Shumilov, and V.G. Giliyazov (1976): Effective loss coefficient variations during PCA. In Auroras and Precipitation of Auroral Particles, Izdatel'stvo Nauka, Leningrad, 104 (In Russian).
- Zmuda, A.J., and T.A. Potemra (1972): Bombardment of the Polar-Cap ionosphere by solar cosmic rays, Review Geophys and Space Phys., 10:981.
- Åbom, C., F.H. Reder, and S. Westerlund (1969): Effects of polar cap absorption events on VLF transmissions, Planet. Space Sci., 17:1329.

Additional references:

- Finger, F.G., M.E. Gelman and R.M. McInturff (1974): High-level circulation studies based on rawinsonde, rocketsonde and satellite observations. In Space Research XIV, Akademie-Verlag, Berlin, 17:
- Laštovička, J. (1977): SID-monitoring at the Panská Ves Observatory, Geofysikalni Sbornik, 23:315.

## VII. SOLAR WEATHER/CLIMATE PREDICTIONS

A WORKING GROUP REPORT by K. Schatten, Chairman, R. Goldberg, J. M. Mitchell, R. Olson, J. Schäfer, S. Silverman, J. Wilcox and G. Williams.

### 1. OVERVIEW

The field of Solar-Activity Weather/Climate offers a promising field for future work. Statistically significant effects will need to be based on physical models and/or the origin of statistical sun-weather climate anomalies elucidated. The fact that much of the public and a good fraction of international scientists feel that there could be solar variability influences upon terrestrial weather and climate, make the subject worthy of study - even if only to disprove any such effects. It is the job of scientists to objectively chase these mysteries of the universe, energetically tackle and illuminate them as best they can. With this in mind, our Working Group on Sun-Weather/Climate studies has investigated both the positive and negative findings in this area, with specific predictions, areas of further study and Group recommendations listed in Sections 5 and 6.

Scientific objectivity suffers when both sides of an issue are either not adequately discussed or the discussions are emotionally dominated. Thus we have attempted to present both sides of the sun-weather controversy objectively. If we have erred in this regard, it is in favor of sun-weather effects, as our group is optimistically inclined in this regard, although we feel we have a healthy skepticism.

The field of sun-weather relations has suffered from a number of problems. These are:

- 1) Numerous reported "effects" which could readily have been disproved on further checking prior to publication;
- 2) A lack of adequate use of statistical tests;
- 3) The use of standard statistical tests not taking proper account of persistence and inherent meteorological quasi-periodicities, both of which tend to overestimate the significance of statistical correlations, regressions, etc.;
- 4) Non-objectivity of scientific work, in which the researcher subconsciously "demands" that either a) an effect must exist and it is only a matter of searching for it until one "finds" it, or b) all solar-weather effects are invalid and will "disappear" given sufficient time and energy to disprove;
- 5) A lack of mechanisms, theories, and hypotheses which can be tested;
- 6) Inadequate data bases for solar, weather, and "intermediary" parameters suited to these studies.

The problems 1 through 3 are reviewed by Pittock, A. B. (1978). Problems 3 and 4 are reviewed in Section 2. Problem 5 is reviewed in Section 3.

Problem 6 is then taken up in Section 4.

To the many sun-weather skeptics, we would like them to consider the long history of science, with its equally long history of groping forward. Particularly during this 100th anniversary year of the birth of Einstein, we might call attention to the development of his Special Theory of Relativity which was able to transform a chaotic set of puzzling and conflicting observations and theories into a holistic and firmly established science. Meteorology is on a firmer footing than the aether theories at the turn of the last century, nevertheless there are some very good indicators which suggest our atmosphere is not completely isolated from solar variability influences; it does not seem probable that another Einstein will emerge and solve the entire problem virtually single-handed.

The sun-weather problem may not be as fundamental relative to physics, but is more complex. To unravel it seems to require as much knowledge of meteorology, solar physics, plasma physics, cosmic rays, chemistry, statistical and advanced mathematics as the researcher can bring to bear on the problem. Problems exist both in theory and experimentation. In this latter field, experimentation generally means "observation" and is limited to what nature chooses to reveal to us through observations of opportunity. These, of course, cannot be repeated in the manner that a physical scientist can repeat experiments in the laboratory.

We observe and work within a relatively short time scale (days, weeks, years) with solar conditions changing on both short and long time scales. This can make the researcher myopic. Recent advances in solar and sun-weather research have benefitted from the far-sightedness of some recent investigators (e.g. Eddy, 1979).

## 2. OBJECTIVITY AND STATISTICAL PROBLEMS

### 2.1 Comments on Existing Sun-Weather Work

Even a cursory look at the literature on sun-weather relationships demonstrates a confusing welter of contradictory results and conclusions. We feel that much of this confusion results from a lack of clarity in the definition and characterization of problems, their context, and the approximate criteria for judging the validity and significance of the results. Proposed sun-weather relationships typically involve a complex web of interacting causalities. Each stage may itself be subject to factors which provide different states for the same inputs, and thus different results. Simple causality, in the sense of one-to-one cause and effect, cannot be expected. The researcher is then obligated, both to himself and his colleagues, to carefully determine and characterize, to the maximum extent possible, all causal links which relate the parameters under study. We recommend that researchers and the colleagues who read and judge their work adhere to the following guidelines:

- 1) State explicitly the hypothesized chain relating the parameters under study. If the study is correlative (based on ad hoc empiricism) only, state this explicitly.



- 2) Specify, for each link in the chain, such intervening or superceding factors that may influence or change the state of the system through the link. Specify whenever possible the values of these factors at the time of the experiment.
- 3) Specify the mechanisms which are hypothesized to be operative for each stage and critically evaluate the certainty with which these mechanisms are known. Include such factors as reaction rates, energy budgets, radiation inputs, and such other factors as may significantly affect the final result.
- 4) Critically evaluate the range of possible results for the entire chain. Evaluate its utility for a) understanding of the physical basis of the phenomenon; b) the testing of hypotheses; and c) predictive purposes.
- 5) Independent analyses of reported effects or mechanisms greatly add to the credibility, particularly if new or differing data are employed.

## 2.2 Guidelines for Future Sun-Weather Work

Because of the complexity of sun-weather relationships, it is unlikely that simple cause-effect experiments may be possible except for some situations. The experimenter will often be forced to rely on masses of data, and thus statistical results. These should be interpreted according to the following guidelines (from A. B. Pittock, 1978):

- 1) Understand the properties of the data, their errors, biases, scatter, autocorrelation, spatial coherence, frequency distribution, and stationarity.
- 2) Choose statistical methods appropriate to both the properties of the data and the purpose of the analysis (e.g., description or prediction).
- 3) Critically examine the statistical significance of the result, making proper allowance for spatial coherence, autocorrelations and smoothing, and data selection.
- 4) Test the result on one or more independent data sets or subsets of the original data.
- 5) Endeavor to derive a physical hypothesis which can be tested on independent data, preferably at some other stage in the hypothesized chain of cause and effect.
- 6) Estimate the practical significance of the result, e.g., the fraction of the relevant total variance which can be predicted or "explained".
- 7) Carefully explain the properties and limitations of the data, the statistical methods used (including data selection and smoothing), and any assumptions, reservations, or doubts.
- 8) Do not overstate the statistical or practical significance of the result.

While statistical methods are likely to continue to be predominant in these studies, we recommend that wherever possible, alternative methods of hypothesis testing be developed and explored. We call attention, in particular, to the Bayesian approach suggested by P. A. Sturrock (1973) for the testing of astrophysical hypotheses. The method, as far as we are aware,

has not been applied or tested in the field of sun-weather relationships, and its potential for achieving consensus there is unknown. We feel, however, that in this field, in which widely divergent results may follow from similar inputs, more effective criteria for achieving consensus than those we now have will have to be found.

Because of the complexity and elusive nature of the many intervening processes, mechanisms, and feedbacks which characterize the Sun-Weather relationship, often coupled with inadequate data bases to illuminate them, the scientific mind searching for confident and neatly wrapped answers may become uncomfortable with this topic. Indeed, the answers may be quite elaborate and "messy", but the urgent need for bright minds to bend themselves in search for them is undiminished.

### 2.3 Further Ideas

The usual statistical approach to searching for a sun-weather relationship is obtained by considering a solar index believed to reflect a certain type of solar influence as an independent parameter, and then finding a subsequent meteorological response. However, an alternative statistical approach is to start with the statistical analysis of weather pattern changes and relate those to one or more indices of variable solar activity.

We feel that sun-climate studies are still in a primitive observational stage of their development. It is premature to attempt the construction of physical models without first, or in parallel, providing a solid observational connection between observed climatic variance patterns and certain indices of variable solar activity that are considered likely to reflect physical changes on the sun of a relevant kind.

Furthermore, the observational base for the construction of prediction models may properly depend upon the application of non-linear, as well as linear, techniques of statistical description and prediction to the identification of spatial and temporal patterns of the fields of climatic response to the selected indices of variable solar activity. The establishment of the reality of any relationships by the demonstration of their predictability can then form a basis for the construction of physical models.

## 3. INTERMEDIARY MECHANISMS BETWEEN THE SUN AND WEATHER

### 3.1 Overview of Intermediary Mechanisms

Several hypotheses for physical mechanisms or links to couple the energetics of solar activity to meteorological responses have been advanced. However, none can be said to be sufficiently complete to be applied to weather or climate prediction. Additional experiments, studies and analyses are required before sun-weather physical concepts can be utilized for sub-predictions. A large number of the physical processes which have been proposed are reviewed in Herman and Goldberg (1978) and Goldberg and Herman (1979). These processes can be divided into two general categories, the middle atmospheric role, and the electrical connection.



### 3.2 Middle Atmospheric Role

The first consideration is the role of the middle atmosphere (15-100 km) as a buffer region, where energy sources from external sources modulate local atmospheric parameters, which in turn affect the transmission, reflection, amplification, etc. of energy inputs both from extraterrestrial and tropospheric drivers. For example, solar particle streams might affect the chemical composition in the upper stratosphere, including the distribution of ozone, leading to changes in temperature and wind fields at this level. This in turn could affect the vertical propagation of planetary waves of tropospheric origin, leading under certain conditions to enhanced reflection or absorption, and, through that, to a modulation of tropospheric pressure patterns and their associated weather and climate conditions. Other middle atmospheric phenomena of potential importance include aerosols and particulates, including noctilucent clouds, nacreous clouds, submicroscopic particulates, etc. Suggestions regarding the influence of those particulates include transport of the neutral atmosphere by drag, radiation balance, scattering of incoming radiation and others.

### 3.3 The Electrical Connection

A second category considers the electrical environment and its relationship to weather and climate phenomena. Until fairly recently, the studies of geoelectricity, geomagnetism, and atmospheric electricity were largely separate endeavors; and, furthermore, unrelated to meteorological phenomena. Now we find that geomagnetic and electric field fluctuations are interlinked, and suggest a path by which extraterrestrial influences on thunderstorms, precipitation, cloud formation, etc. are possible. To complicate this issue, man-made disturbances created by power lines, radioactive waste, etc., could conceivably impact on the atmospheric electrical circuit. Several theories now contend that modulations of solar particle radiation and of galactic cosmic rays by solar events can affect the atmospheric electric field, on either a global or local scale, which may cause enhancement of lightning activity and electric fields within thunderstorms. A possible scenario to bridge the gap to meteorology might then occur as follows: modification of the tropospheric electric field affects charge separation in thunderclouds (especially in tropical convective storms); this leads to enhanced electric field strengths within clouds. A result is to modify the droplet coalescence efficiency and hence the drop size distribution in clouds which might affect the vertical distribution of liquid water (and so the latent heat distribution) within the clouds. This could alter the convective instability and vertical development of clouds and lead to a vertical flux of horizontal momentum of the planetary scale windfield. Potentially, changes in weather and climate at all latitudes could result. The above mechanisms are illustrative of a number of proposed mechanisms for which a recommended approach of investigation and testing is possible.

## 4. SOLAR AND TERRESTRIAL PARAMETERS

When contemplating which solar parameters might be significant to weather and climate, and which aspects of weather and climate might



reasonably be expected to respond to solar variations, we need to keep several things in mind:

- 1) We are not certain of the form(s) of energetic flux changes on the sun that may be of meteorological significance, or of the time scales of such changes. What we do know is that the sun is capable of some kinds of changes of energetic output (and magnetic effects at the distance of the earth), but our knowledge concerning these changes is both qualitatively and quantitatively incomplete.
- 2) The Earth's weather and climate are highly variable. It is a distinct possibility that the total variability can be divided into two parts: a) the inherent or intrinsic atmospheric component and b) the external component caused by such outside forces as solar and lunar effects.
- 3) Weather prediction models, and climate models, are a means of quantitatively assessing the inherent variations of weather and climate by attempting to synthesize all of the important internal physical and dynamical processes in the atmosphere that contribute to weather and climate variations. Because these models do not allow for solar effects, studies of the comparative skill with which such models "predict" climate behavior during quiet and during disturbed solar conditions can provide useful insights into the impact of solar events or changes on weather phenomena and on weather (or climate) predictions as such. These insights are a kind of "back door" variety.
- 4) Systematic research to identify solar impacts on weather and climate is greatly handicapped by conditions 1) and 2) above. Since weather and climate models leave out many physical influences in addition to solar influences, opportunity 3) is not sufficient as a research strategy. Some suggestions are the following:

Terrestrial weather and/or climate could, in principle, be influenced by variations in any of a variety of forms of energy emitted by the sun in the direction of the Earth. Among these forms of energy are: (a) the total solar irradiance ("solar constant"), (b) spectral solar irradiance, especially UV, (c) solar wind, (d) high-speed particle streams, and (3) x-rays. A fuller list of possibly useful parameters is given at the end of this section. The fuller list consists of indices which are not all directly energy related but may be indicative of energy fluxes. It is imperative that a means be found and implemented to monitor on a continuous basis these various solar energy fluxes, together with the interplanetary magnetic field (IMF), to establish the nature and extent of variations in each together with the time scales of the variations. The relevance to weather and/or climate should be explored in two ways:

1. by studying atmospheric events that appear to follow solar flux changes, and for which a physical connection is suspected from either theoretical or empirical evidence;
2. by relating the solar flux changes to conventional solar indices, if possible, to clarify which of a variety of available solar indices might serve as proxy measures of the longer-term flux changes in earlier periods of time, as needed for the clarification of longer-term sun-climate relationships. Examples of such solar indices might be sunspot numbers, sunspot areas, sunspot umbra-penumbra ratios, solar flare indices, plage areas,

Table 1. Parameters of Potential Value for Study of  
Sun-Weather Conditions

## I. Solar Parameters

- |                                   |   |
|-----------------------------------|---|
| 1. Sunspot Number                 | 5. Solar Wind (Interplanetary)              |
| 2. Active Regions                 | a. High Speed Streams                       |
| a. Sunspot Area                   | b. Particle Composition                     |
| Umbra        }                    | c. Magnetic Field                           |
| Penumbra    } Ratio               | d. Energetic Particles (Solar and Galactic) |
| b. Plage                          |   |
| 3. Solar Flares                   | 6. Solar Constant                           |
| a. Location                       | a. Diameter                                 |
| b. Number                         | b. Energy (luminosity)                      |
| c. Size            }              | c. Temperature                              |
| d. Brightness    } Classification | d. Periodicities                            |
| e. Cumulative Flare Index         | e. Variations                               |
| 4. Electromagnetic Radiation      | 7. Solar Magnetic Fields:                   |
| a. X-ray                          | a. Active Regions                           |
| b. UV                             | b. Coronal Holes                            |
| c. Radio                          | c. General Fields                           |

## II. Terrestrial Parameters

### A. Meteorological

Physical and Chemical

Parameters may be subdivided as being atmospheric parameters above 100 km, below 100 km, or oceanic parameters.

Global parameters are of prime importance, and among those are:

1. Dynamics
  - a. Planetary Waves
  - b. Position of Jet Stream
  - c. Vorticity Area Index (VAI)
2. Energy Balance, Radiation Budget
3. Stratospheric Chemistry

### B. Non-Meteorological

1. Greenwich Delta "D"
2. Geomagnetic Indexes:  
    AE, Ap, Kp, aa, and Dst
3. Aurorae, Airglow
4. Atmospheric Electrical Circuit



solar magnetic field indices, or even geomagnetic indices known to be related to solar conditions such as coronal holes. Spectrally broadband changes of energy flux (e.g. total solar "constant" changes) would be expected to alter many characteristics of the Earth's weather and climate system on a global scale. Among these characteristics are: (a) planetary heat budget, (b) planetary wave regime (both standing waves and traveling Rossby waves), (c) polar vortex and jet stream pattern, (d) low-latitude Hadley circulation regime, and (e) other mid- and high-latitude features including blocking patterns, stratospheric warming events, and "eigen modes" of circulation patterns as revealed from empirical studies. A further list is given at the end of this section.

Spectrally narrowband flux changes, and/or particle flux changes, or possible IMF effects, would not necessarily affect weather and climate worldwide but perhaps specifically in the higher latitudes. Changes in all the weather and climate characteristics listed above could, however, be investigated in connection with this category of solar effects as well. Special attention may need to be given to the following areas: (a) thunderstorm activity; (b) auroral-zone atmospheric conditions; (c) stratospheric ozone changes (and any other photochemical changes that might be capable of altering either the earth's heat budget or stratospheric circulation); and (d) the state of (geographically or altitudinally) selective regions of the atmosphere which may be particularly receptive to solar activity influences.

In brief, weather and climate parameters to be investigated in connection with the assessment of solar influences should be those which a) are more-or-less global in scale and representation, and b) as much as possible reflect physical conditions of the atmosphere or climate system as a whole (in contrast to local weather and climate statistics). The reasoning behind this is essentially that, although local weather conditions would be affected by a real sun-weather or climate relation, solar forcing of the climate system would generally be global in scale and not local.

## 5. SPECIFIC PREDICTIONS AND FURTHER STUDIES

Despite the feeling that sun-weather research might not be ready for predictions, the working group attempted to assess how far one might be willing to go toward formulation of a meteorological prediction based on presently existing sun-weather research. It was agreed from the outset that any prediction we might attempt should be treated as a purely academic experiment and that minority opinions concerning its credibility, or lack thereof, are to be encouraged as part of our report.

### 5.1 A Sun-Weather (Climate) Prediction--The 22 Year Drought Cycle

As a result of our discussions based on the work of Mitchell et al. (1977), we have chosen to focus on the possibility of long-term prediction of drought in the western U.S. Such a prediction is to be regarded as tentative and only as an academic prediction to be tested against future developments. We want to stress that this prediction is based on the extrapolation of a statistical apparent in historical data and not at all on any understanding



of the physical processes which might be involved. Implicit in this prediction is the assumption that solar variability will continue in the future to exhibit past behavior in terms of the Hale sunspot cycle, etc.

Figure 1 can be interpreted as a prediction of the risk of drought of different areal extents occurring in the Western United States during the current Hale cycle. 1976 may be taken as the zeroth year after sunspot minimum (abscissa). In the analysis leading to this diagram, the United States, west of the Mississippi River, has been divided into 40 areas. The four diagrams, based on different groupings of tree ring data and intensities of drought, indicate as a function of phase of the Hale cycle the risk in a particular year that at least 0, 10, 20 or 30 of the 40 areas have a Palmer Drought Severity Index (PDSI) value greater than (or less than) a given threshold value.

The implication of this figure is that, at present, the risk of severe drought is relatively high but decreasing. Note that no information is given about which of the 40 areas might participate in the lowering risk of drought. It must be stressed that we are only forecasting drought risk and that the drought variations with solar cycle phase represent only a fraction of the total drought variance; we can make no definite statement concerning the presence or absence of a drought in a given year. An increased risk of drought in the western U.S. could be expected near or following the end of solar cycle 22. The effects depicted may be strictly periodic but in any case appear to be phase locked to the Hale cycle; if the current Hale cycle turns out to be short, it may be necessary to adjust the time line of the prediction accordingly. Also, the effect on droughts tends to break down during weak cycles so that although the prediction relationship has a high statistical significance it cannot be counted on to apply with equal force to all solar cycles. We emphasize that this prediction applies only to the Western U.S. Also, it should be emphasized that since the PDSI integrates moisture down to the water table, it is more of a measure of hydrological than agricultural drought and so its applicability to agriculture is not entirely straightforward.

Inasmuch as no causal mechanism for the above statistical effect has been clearly established, we believe that studies of causal factors should be stressed. Conceivably, coronal hole activity, the sun's polar magnetic field strength, or possibly, recurrences of solar wind streams (see Sargent, 1979), are physically related to the drought index (mainly rainfall) variations.

## 5.2 Short-Time Scale Effects

### 5.2.1 Vorticity Area Index and Related Effects

The early work of Markson (1969) first suggested solar wind parameters may relate to terrestrial atmospheric electrical properties. Recent evidence was discussed for a specific link between solar wind parameters (including IMF and CRs) and tropospheric circulation. Wilcox et al. (1979) show that low pressure (300 m bar) troughs (cyclones) that are near 180°W longitude when the interplanetary magnetic field is directed away from the sun are, on the average, significantly larger than when the field is toward the sun. The

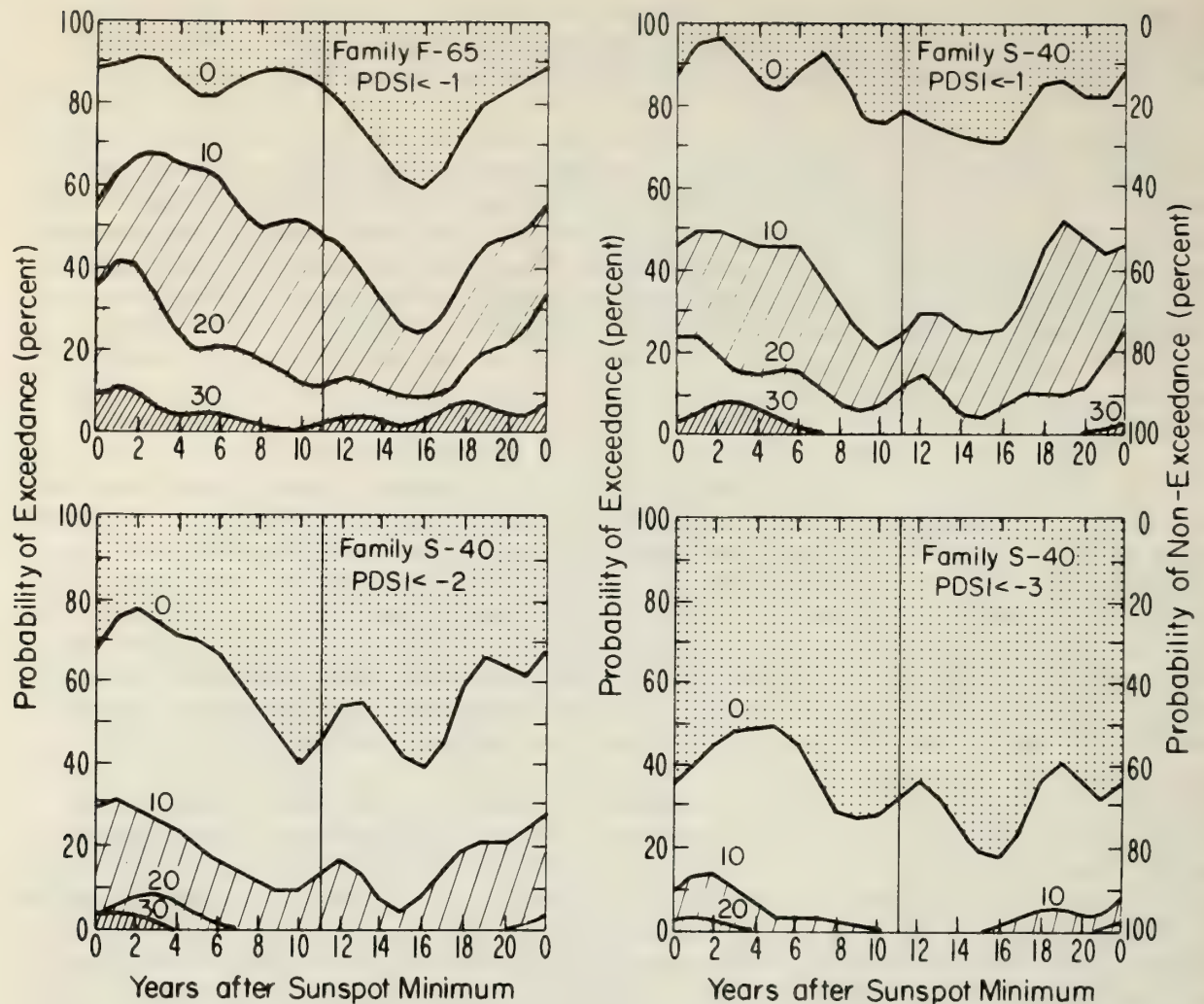


Figure 1. Probability of exceedance of different areal extents of drought (interior numbers, indicating number of regions out of a total of 40) as a function of the phase of the Hale sunspot cycle (adjusted to its average length of about 22 years). Based on data for period 1700 - 1962 A.D., for each of four tree-ring "families" and Palmer Drought Severity Index (PDSI) limits indicated. All curves have been smoothed as a function of time by a binomial moving average with weights 1-4-6-4-1 over 16.

difference in area persists during a 5-day interval in which the troughs move from  $180^{\circ}\text{W}$  to the North American Continent. This relation persists during most of the winters from 1951 to 1973. This is a link between the direction of the IMF and the area of individual troughs. Variations in the solar constant or the sunspot number are not involved.

Another such specific link is a report that a sub-group of particularly active sector boundary (SB) transits gave a VAI sun-weather effect twice as large as usual. This relates to an apparent "dip" in the VAI index one day after a sector boundary crossing. The active boundary transits were defined from a list published in Solar Physics by Svestka et al., of transits



accompanied by streams of MEV protons observed in the interplanetary medium. Some clues to physical mechanisms may be obtained as follows. The sun-weather effect was twice as large for the active boundary transits, but was clearly distinguishable for the other transits. The IMF magnitude and also geomagnetic activity had a large increase after the active boundary transits but no significant change after the other transits. The galactic CR intensity had a large (1%) decrease after the active transits but no significant change after the other transits. This suggests that changes in IMF magnitude, geomagnetic activity and CR flux are not directly related to this particular sun-weather effect.

The above suggests that changes in the amplitude of solar and interplanetary parameters may influence the amplitude of tropospheric circulation responses. It also seems likely that the state of tropospheric circulation can determine whether it will be responsive to a small nudge from the sun. In particular, it has been reported (Wilcox and Scherrer, 1979) that the amplitude of a sun-weather effect is proportional to the variance in tropospheric circulation intensity, as was discovered by Hines and Halevy (1977). When this effect is properly accounted for, the influence of the sector structure on the Vorticity Area Index has been remarkably constant from 1963 to the present.

A certain amount of discussion centered on the interpretation of this result; Williams and Gerety (1978) call attention to recent changes of this effect. Further, the result of another recent unpublished result suggests that the amplitude of this effect is not dependent on the mean value of the VAI prior to a SB. Neither of these results explains the seasonal dependence of the effect. It was agreed that allowance for the initial state of tropospheric circulation is an important consideration in sun-weather investigation.

One possibility exists concerning the new results indicating that the VAI effect is related to a decrease in the kinetic energy in wavenumber 5. This result may have important implications for identifying mechanisms.

Thus, on the shorter time scales, in view of the concerns related to statistical problems, the working group did not feel able to make any predictions regarding the future behavior of the VAI or other atmospheric parameters in relation to solar or interplanetary features. It is hoped that future studies within the next few years will clarify ideas in these areas and that predictions will become possible.

### 5.2.2 The Electrical Connection: Thunderstorm and Lightning

The working group felt that the results of studies in the electrical area looked very promising for future predictions, but were not sufficiently well understood to lend themselves to predictions at present. Solar flares and interplanetary sector boundaries were parameters showing signs of being related to thunderstorms. Future satellite and ground-based observations should provide sufficient new data to test suggested relations or devise more permanent ones within a relatively few years.



### 5.3 Disappearance or New Parameter—Explanation of Effects

In connection with the statistical problems as outlined in section 2, it was noted that many reported sun-weather relations have had a tendency to either "disappear" or require additional post facto parameters to explain more recent behavior of these effects. A minority opinion felt that most short time-scale sun-weather relations may not be real and the above situation is an indication of this. The majority realized these problems existed, but were optimistic concerning the reality of real, short time-scale sun-weather relations.

As there is no well established academic discipline in sun-weather studies, it is not surprising to find research in this field of mixed quality. Thus, much work done in solar-weather research may in time be seen to be invalid; however, good research in the field is possible, has been done, and hopefully as time goes on will rise to even higher levels of quality. The existence of poor work does not mean that the field is unworthy of serious inquiry or that it is incapable of providing an opportunity for high-quality and fruitful studies.

## 6. RECOMMENDATIONS

1. We endorse an active experimental program to define, develop and conduct coordinated experiments to investigate various aspects of the more promising sun-weather and sun-climate mechanisms that have been proposed. In particular, emphasis should be placed on establishing the response of the middle and lower atmosphere to changing radiation of various forms directly related to solar active phenomena. The natural variability of the atmosphere under such influences is not well understood at this time. In the same connection, experiments should be developed to investigate the electrical properties of the atmosphere under different conditions of solar activity. Response of thunderstorms, and their internal properties, should also be investigated.
2. The role of ozone, including its global distribution, its possible response to solar and geomagnetic changes, and its effects on atmospheric heating should also be studied. The ozone data currently provided by Nimbus IV and VII offer a valuable basic data set for such studies.
3. Theoretical modeling of atmosphere response to changes in ozone is required to evaluate the role that stratospheric ozone may play in modifying tropospheric weather and climate through circulation and heat budget.
4. Techniques to measure the global electrical circuit should be developed. Elucidation of the response of the global circuit to solar activity is an important first step to assessing several current theories regarding possible meteorological effects of solar activity through this electrical connection.
5. Observations of lightning and thunderstorm frequency in both the northern and southern hemispheres are required to evaluate possible correlations with solar activity phenomena.

6. A pilot forecast program should be defined and developed to test the possibility of using solar activity related indices in meteorological forecasts.
7. Studies of the recurrence patterns of the aa indices, the sun's polar magnetic field, and other solar and solar wind parameters should be furthered to ascertain the causal mechanism involved in the 22-year cycle found in drought in the western U.S. and in some other climatic conditions.
8. Assessment of drought cycles in localities other than the western U.S.; a possible global pattern should be undertaken as available tree-ring and other "proxy" climatic data permit.
9. It is recommended that the hourly values of proton flux measurements at 1 AU, that have been measured by various satellites since about 1965, be reduced to daily and hourly values for the energy ranges 1-10 MeV and more than 10 MeV. Current and proposed measurements of the X-ray, EUV and UV fluxes from the sun as well as solar wind plasma, and magnetic field properties on a daily basis are also essential.
10. With reference to the need for a continuous monitoring of the solar constant, we would have thought it unnecessary to stress the importance of this, but for emphasis, we do so here. Likewise, all available means of reconstructing the history of solar luminosity, in past years, centuries, and millennia, should be exploited as a matter of urgency in helping to unravel the causes of past climatic changes. Similarly, great emphasis should be placed on future short term (days) as well as long term (years) solar luminosity observations.

## 7. ACKNOWLEDGMENTS

We wish to acknowledge the contributions to the group's thinking and efforts by the following group members: A. Belmont, J. Bryson, R. Head, M. Lethbridge, E. Prior, B. Quate, and H. Willett (who were present in body), and J. Herman, M. Larsen, and H. Volland (who were present in spirit). We further thank R. Donnelly and his staff for their herculean efforts in providing the group with important background information and with fostering the arrangements necessary to interact with several other working groups, as needed for us to develop this report.

## 8. SOLAR-WEATHER/CLIMATE WORKING GROUP

K. Schatten (Chairman)  
Astronomy Department  
Boston University  
Boston, Mass. 02215\*

\*Now at NASA Goddard Space Flight Center

R. Goldberg  
Laboratory for Planetary Atmospheres  
NASA Goddard Space Flight Center  
Greenbelt, Maryland 20771

J.M. Mitchell  
Environmental Data and Information Ser.  
NOAA  
Silver Spring, Maryland 20910

R. Olson  
Aspen Institute for Humanistic  
Studies  
Boulder, Colorado 80302

J. Schäfer  
Radio Astro. Inst.  
University of Bonn  
5300 Bonn  
Federal Republic of Germany

S. Silverman  
Optical Physics Division  
AFGL/L.G. Hanscom Field  
Bedford, Massachusetts 01730

J. Wilcox  
Institute for Plasma Research  
Stanford University  
Stanford, California 94305

G. Williams  
Department of Physics  
The University  
Southampton, SO9 5 NH  
Great Britain

## 9. REFERENCES

- Eddy, J. (1979): The New Solar Physics, AAAS Symposium #17, Westview Pr., Boulder, CO.
- Goldberg, R. A., and J. R. Herman (1979): these proceedings.
- Herman, J. R., and R. A. Goldberg (1978): Sun, Weather, and Climate, NASA, SP426.
- Hines, C. O., and I. Halevy (1977): Nature, J. Atmos. Sci., 39, 382.
- Markson, R. (1969): Electrical Processes and Problems in the Stratosphere and Mesosphere, IAGA, Vol. 84, Madrid, Spain.
- Mitchell, J. M. Jr., C. W. Stockton, and D. M. Meko (1977): Drought Cycles in the United States and Their Relation to Sunspot Cycle Since 1700 A.D., EOS, 58, 694.
- Pittock, A. B. (1978): Rev. Geophys. and Space Phys., 16, 400.
- Sargent, H. H. III (1979): Solar Terrestrial Influences on Weather and Climate, ed. McCormac, D. Reidel.
- Sturrock, P. A. (1973): Astrophys. J., 182, 569.
- Wilcox, J. M., P. B. Duffy, K. H. Schatten, L. Svalgaard, P. H. Scherrer, W. O. Roberts and R. H. Olson (1979): Science, 204, 60.
- Wilcox, J. M., and P. H. Scherrer (1979): Nature, to appear.
- Williams, R. G., and E. J. Gerety (1978): Nature, 275, 200.



# COUPLING PROCESSES RELATED TO THE SUN-WEATHER PROBLEM\*

Richard A. Goldberg  
Laboratory for Atmospheric Sciences  
Goddard Space Flight Center  
Greenbelt, Maryland 20771 U.S.A.

and

John R. Herman  
Radio Sciences Company  
624 Tulane Avenue  
Melbourne, Florida 32901 U.S.A.

Physical mechanisms for coupling the energetics of solar activity to meteorological responses are reviewed. Although several hypotheses have been advanced, none can be said to be sufficiently complete to be applied to weather or climate prediction. Solar activity indicators potentially useful for forecasting are identified, including sunspots, solar flares, and magnetic sector boundary crossings. Additional experiments, studies and analyses are required before sun-weather concepts can be utilized for predicting meteorological responses.

## 1. INTRODUCTION

The application of sun-weather processes to weather and climate prediction presupposes one or more observable solar activity precursors which will be followed by known meteorological changes within a definable time lag. The sequences of events must be repeatable, and the number of observable variables defining them should be not only finite but reasonably small. In order to specify the required variables the physical processes underlying the sequences should be understood.

Within this framework it appears that the present knowledge of sun-weather relationships is inadequate for practical application to forecasting. However, recent findings and reassessments of older sun-weather studies suggest that at least some meteorological changes are associated with solar activity, although it is not yet clear whether the association is a statistical accident or has a cause rooted in physical coupling mechanisms. Thus, sun-weather relationships offer promise for application to forecasting, but additional study and understanding is required before that potential can be realized.

The purpose of this paper is therefore to review the present status of the field. There being a great deal more interest in the field than there is work actually being performed, a significant fraction of the subject literature already consists of reviews (c.f.: King, 1975; Wilcox, 1975; Meadows, 1975; Pittock, 1978; Herman and Goldberg, 1978b), so here possible mechanisms are emphasized, and an attempt is made to suggest areas for further progress and study. In the present context, solar

\*Invited Review for the International Solar-Terrestrial Predictions Workshop, Boulder, Colorado, April 23-27, 1979.

activity encompasses not only transient solar and geomagnetic phenomena, but also modulations of the galactic cosmic ray flux reaching Earth, and solar rotation effects including solar magnetic sector boundary crossings.

The nature of the sun-weather problem is summarized in Section 2, followed by a brief synopsis of solar radiation sources (Section 3). Coupling processes are addressed in Section 4. Included are effects on ozone leading to possible stratospheric heating and subsequent alteration of atmospheric circulation; the role of aerosols and particulates in modulating atmospheric transmission; and atmospheric electrical processes. Solar activity indicators as precursors for weather and climate changes potentially useful for prediction (pending additional studies) are discussed in Section 5.

## 2. NATURE OF THE PROBLEM

Improvements in the forecastability of weather and climate by incorporation of the variance introduced by solar activity will be made easier if causal mechanisms are understood. A key factor in establishing such mechanisms rests in the recognition that different components of the Sun's energy output interacts with the Earth's atmosphere in different ways. No less important is the fact that the various components exhibit different degrees of time variability, so the atmospheric response to "solar activity" can exhibit a variety of forms and time lags. Thus, an orderly progression toward achieving improvements in forecastability would be to determine: (1) the variability of each solar output component; (2) the atmospheric response to each component; (3) the amount of variance in the meteorological parameter attributable to the solar signal; (4) the influence of each parameter on weather or climate changes; (5) methods to incorporate results into prediction programs.

Figure 1 (Sechrist, 1979) helps exemplify the complexity and nature of the problems relating to solar terrestrial coupling. Solar transient and variable emissions reach the Earth in the form of both electromagnetic and corpuscular radiations. These radiations in turn affect the chemistry and physical properties of the upper, middle and/or lower atmosphere. Timing for particle arrivals and subsequent magnetic field controls are also important. Electromagnetic radiations travel to the Earth in a few minutes at the velocity of light and affect all sunlit regions; corpuscular radiations arrive from a few hours to a few days following a solar event and are primarily focused to high latitude regions. Other effects include galactic cosmic rays, which do not originate from the Sun but are modulated by short and long-term (solar) variability. Solar magnetic field effects induced by solar rotation may also couple to the Earth electromagnetic structure. Finally, meteoric ions may be modulated by ionization sources subject to solar variability.

Figure 1 concentrates on processes involving the passage of signals, energy, and other forms of coupling from the upper to lower atmosphere. There is already a large body of evidence demonstrating the reverse influence i.e. tropospheric meteorological disturbances on the upper atmosphere and magnetosphere. These add to the difficulty of analyzing cause and effect, and must be contended with in suitable coupling mechanism concepts.

For example, Gherzi (1950) and Bauer (1958) demonstrated ionospheric F region ionization enhancements above hurricane passages. Apparent ionospheric responses to cold fronts were also noted by Gherzi (1950), Bauer (1957), and Arendt and Frisby (1968). To explain the observational results, Bauer (1957) hypothesized a dynamic coupling process between the troposphere and ionosphere.

Acoustic waves generated by severe local storms have been detected in the F region by Georges (1968), Baker and Davies (1969), and Hung et al. (1975). The waves, whose periods in the ionosphere are in the range 2-5 minutes, are detected by CW doppler techniques. Most recently, Hines and Halevy (1977) have proposed a gravity wave feedback mechanism, whereby energy produced in the troposphere by meteorological phenomena propagates upward via gravity waves. Under certain



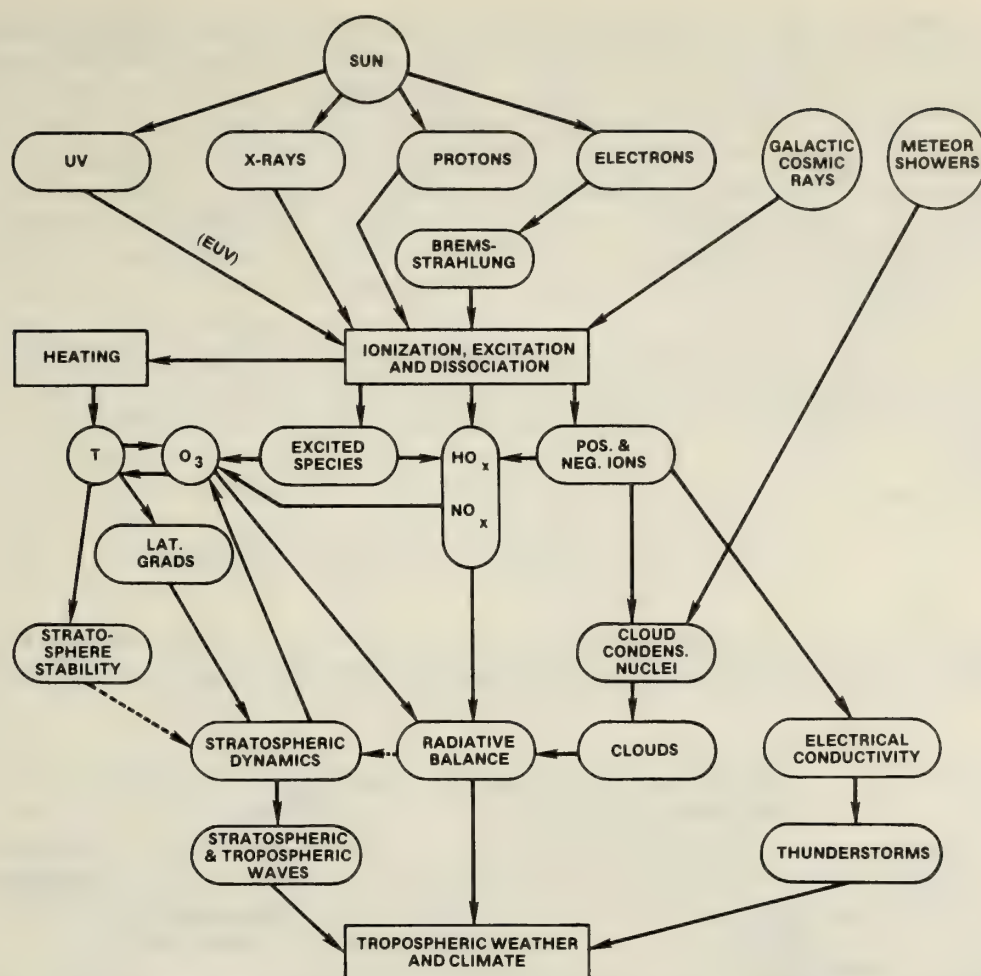


Figure 1. Schematic depiction of extraterrestrial radiation sources modulated by solar activity, and their subsequent effects and interactions with various parts of the middle and lower atmosphere. (after Sechrist, 1979).

assumed conditions which could be modulated by solar activity, the upper atmosphere will reflect the waves downward to interfere constructively or destructively.

In addition, there is evidence that lightning can induce effects in the middle atmosphere and above. For example, whistlers, which are VLF waves thought to be induced by lightning discharges, are guided along the Earth's magnetic field between conjugate points for path lengths exceeding 30,000 to 50,000 km (e.g. Davies, 1965). Recently, Bering et al. (1979) have measured microbursts of electron precipitation from the magnetosphere, stimulated by VLF spherics generated by lightning in a thunderstorm cell. Since the precipitating particles can ionize and dissociate molecules in the upper atmosphere, a lightning induced mechanism for perturbation of the upper atmosphere is created.

A direct effect on stratospheric ozone has also been suggested by Dubin and Zipf (1977). They have determined enhanced production rates for  $N_2O$  under lightning discharge conditions, and offer the suggestion that large quantities of  $N_2O$ ,  $NO_x$ , and other trace species are produced within the tropical zone, where thunderstorm activity is highest. Injection of such species into the lower stratosphere, particularly within large convective thunderstorm cells that penetrate the tropopause, could significantly deplete stratospheric ozone in time (Noxon, 1976, Chameides et al., 1977).

Man induced effects in the troposphere also affect the upper domains. The most famous topic under discussion in recent years has been depletions in stratospheric ozone induced by



hydrofluorocarbons, which occur in many products such as aerosol sprays. SST transports have also been considered as significant candidates for ozone modification in the stratosphere. Man may also affect the upper atmospheric environment in electrodynamic ways. For example, Park and Helliwell (1978) have shown correlative data to suggest that power lines, particularly at high latitudes, generate VLF waves which propagate to the outer magnetosphere. There they can induce particle precipitation which affects ionization, etc. within the lower thermosphere (Figure 2). Finally, Boeck (1978) has considered the gradual increase in the troposphere of radioactive Krypton 85 discarded during nuclear power plant operation. He suggests that this is rapidly becoming the dominant ionization source in the lower troposphere, and will soon have a profound effect on the global electrical circuit.

The above provides both transient and long term examples of effects that must be considered in any studies involved with the determination of atmospheric coupling processes. The evidence for upward coupling is established. We must now determine if downward coupling can play as significant a role.

### 3. RADIATION SOURCES – A BRIEF SYNOPSIS

Solar electromagnetic radiations are predominantly in the visible, and affect the lower troposphere. There is some evidence (Heath and Thekaekara, 1977) that solar UV, especially toward the

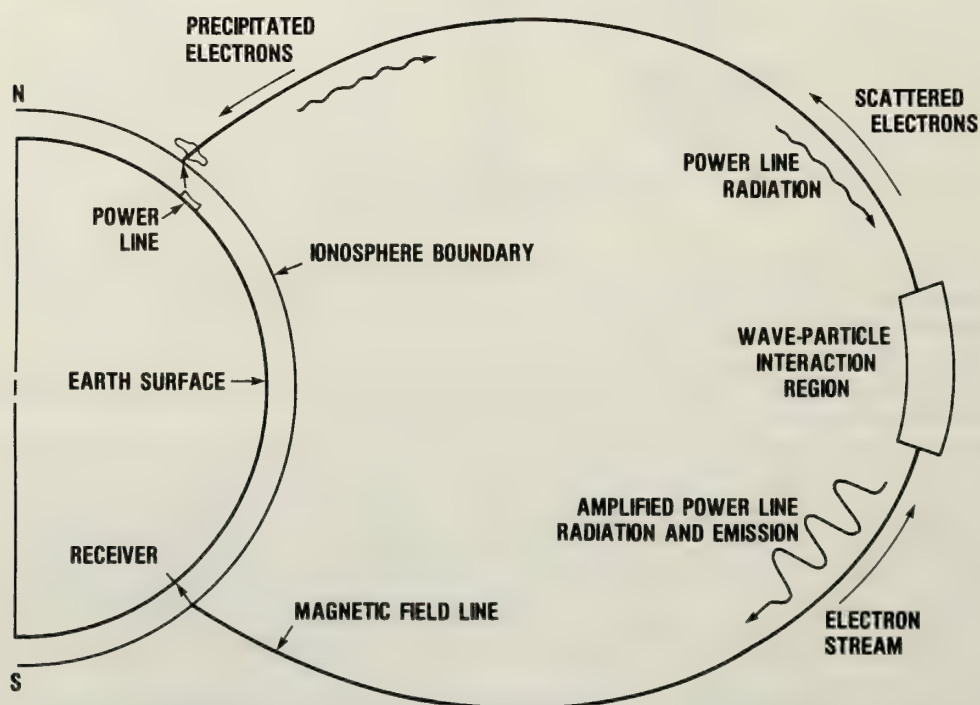


Figure 2. A schematic diagram illustrating the effects of power line radiation (PLR). Harmonic radiation from power lines ( $\sim$  several kHz) is guided to the equatorial magnetosphere where it interacts with counterstreaming energetic electrons. The wave particle interaction amplifies the PLR, triggering emissions of order  $10^3$  greater than the input level. Scattered electrons precipitate into the upper atmosphere with estimated energy fluxes  $10^6$  or more times the input wave power (after Park and Helliwell, 1978).

more energetic portion of the spectrum, shows a solar cycle dependence, but this has not been observed for the total solar output (solar constant). Solar flares and other active phenomena exhibit enhanced emissions of UV and X-rays which cause ozone production and ionization respectively, but such emissions are short lived and do not cause sustained effects.

Solar active events cause the deposition of energetic corpuscular streams into the Earth's atmosphere at high latitudes, where geomagnetic shielding effects are weakest. These energetic particles can directly or indirectly affect the middle atmosphere, depending on their type, energy, and flux. They include solar protons which produce polar cap radio absorption events (PCA), and relativistic electron precipitation events (REP) whereby relativistic electrons created by plasma interactions in the magnetosphere precipitate to stratospheric depths. Furthermore, most high latitude geomagnetic disturbances lead to energetic electron precipitations which cause x-ray aurorae by the bremsstrahlung process near 100 km. The converted energy can reach to stratospheric depths, and can dominate over the effects of galactic cosmic rays to altitudes below 40 km. Finally, galactic cosmic rays are modulated by the individual geomagnetic events and by sunspot cycle, and these have bearing on the ozone distributions in the mesosphere, stratosphere, and troposphere. Characteristic penetration depths for the various corpuscular radiations as a function of energy are illustrated in Figure 3 (Thorne, 1977).

#### 4. COUPLING PROCESSES AND MECHANISMS

In the search for identification of physical processes to explain the statistical correlations of sun/weather and climate, the stratosphere and mesosphere must be regarded as an important atmospheric domain within which coupling might occur. Ozone plays a major role in the radiative and photochemical processes of this region, and therefore must be considered a key parameter for study. Because the variability of ozone may be influenced by ion-neutral chemistry induced by energetic charged particles, it is appropriate to consider such effects here. Aerosols and particulates must also be considered here, since some variation in these constituents can also be related to solar activity. Noctilucent and nacreous clouds, as well as sub-macroscopic particles, have been offered as modifiers of tropospheric weather and climate. Hence, this category considers the middle atmosphere as a buffer region, within which energy undergoes reflection, transmission, filtration, amplification, absorption, and/or some other form of conversion.

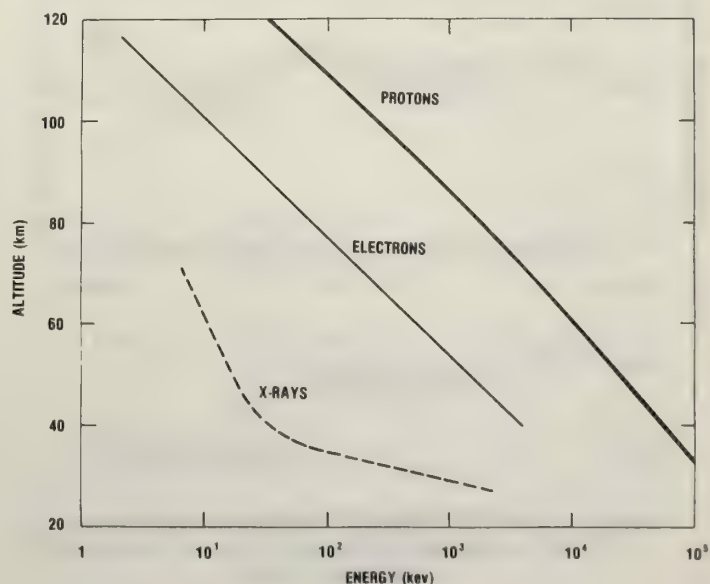


Figure 3. The nominal penetration depth for energetic electrons, protons, and X-rays normally incident at the top of the earth's atmosphere (after Thorne, 1977).



Second, there is newly emerging experimental evidence that atmospheric electric fields, especially at high latitudes and in the upper atmosphere, are responsive to solar and geomagnetic phenomena. The corpuscular radiations previously discussed are known to strongly enhance the local atmospheric electrical conductivity and ionization at stratospheric and mesospheric altitudes. Several new theories have now postulated that some electric field perturbations are regulated by solar activity and may be responsible for the observed statistical correlations between solar activity and thunderstorms.

Both categories contain established results relating to middle atmospheric and tropospheric response to the radiation sources. However, extrapolations to short-term (weather) or long-term (climate) affects in the troposphere all suffer from one or more speculative components. The electrical connection has made more progress toward a full explanation, because it may bypass the stratosphere as an intermediary, and because it causes instantaneous reactions.

#### 4.1 Middle Atmospheric Ozone Responses to Solar Activity

To pass from the upper atmosphere to the troposphere, energy may undergo one or more intermediate changes, possibly affecting relevant meteorological parameters. Recent observations of stratospheric variability in connection with particle injection events have already suggested some possible candidates. A depletion of ozone during a PCA event was first observed in the mesosphere by Weeks et al. (1972) and later interpreted by Swider and Keneshea (1973) to be caused by enhanced  $\text{HO}_x$  production during this event. More recently, a 20 percent depletion above 4 mb was observed during the great PCA event of August 1972, using ozone data from the BUUV experiment on Nimbus IV (Figure 4, Heath, et al. 1977). Most remarkable is the persistence, whereby the  $\text{O}_3$

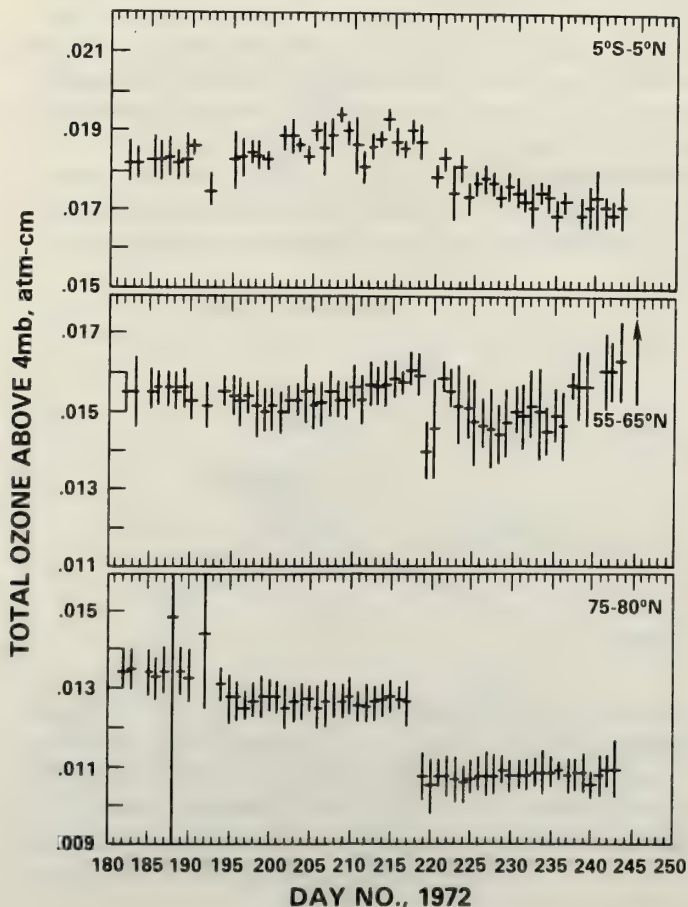


Figure 4. Daily zonally averaged total ozone (Nimbus 4 BUUV experiment) above 4 mb pressure level (about 38 km) for equatorial (top panel), middle (middle), and high (bottom) latitudes during July and August 1972.

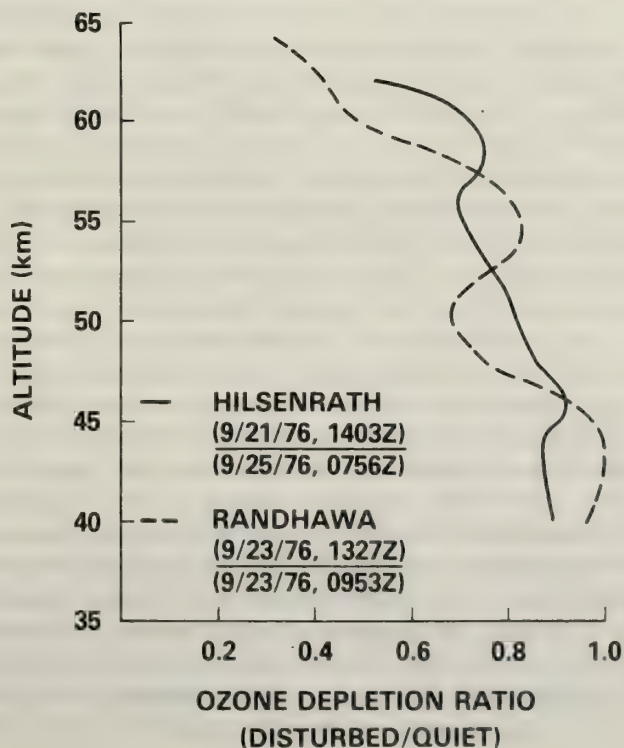


depletion was maintained long after the termination of the responsible source. The high latitude depletions observed were explained by production of odd nitrogen through proton bombardment as a loss agent. It has also been argued by Thorne (1977) that REP events could cause similar magnitude changes when the cumulative effect of these more frequent events are combined. Rocket data evidence for ozone depletions during auroral x-ray events has been reported by Goldberg (1979). Preliminary findings showed a 25% depletion of ozone above 1 mb, cf. Fig. 5, (Hilsenrath, pvt. communication 1979) following each of two nighttime x-ray auroral substorms. These results have been verified with two additional comparisons during March 1978. The observed maximum in x-ray energy deposition occurred in the region of ozone depletion. However, x-ray and particle data showed that nighttime relativistic electrons were also present in adequate quantities to override the x-ray energy deposition above 55 km by as much as a factor of 100. Nighttime relativistic electrons have also been detected on a regular basis during auroral storm conditions by satellite (cf. Reagan, 1977). This newly measured nighttime radiation may help account for the observed ozone depletion, which is too large to be explained by conventional means from the x-rays alone.

Long-term response of stratospheric ozone over a solar cycle (e.g. Angell and Korshover, 1973, 1975) has also been seen from statistical studies based on data from the global network of ground based Dobson stations. If true, this longer response may be more indicative of solar UV variability over a solar cycle, as indicated in the work of Heath and Thekaekara (1977), since solar UV is the primary control for atmospheric ozone production. The results are controversial, because of absolute and intercalibration errors between stations, gaps in data acquisitions and a host of additional problems. The conflicts may further be explained by the fact that total ozone (Dobson) measurements are dominated by the maximum density of ozone found below about 25 km altitude. In this height regime the residence time of atmospheric ozone is about 2 years, and its distribution and concentration are controlled by transport processes. The effects of production and loss processes related to solar variability are thus obscured in total ozone data.

### AUOROZONE I, POKER FLAT, ALASKA, SEPTEMBER 1976

Figure 5. Ratio of nighttime vertical ozone distributions comparing post-auroral disturbance values to quiet time profiles preceding each event. The ratios were calculated from four independent rocket chemiluminescent detector soundings from Poker Flat, Alaska at the times shown. One comparison (Randhawa technique) was obtained on a single night. The second comparison (Hilsenrath technique) was obtained by comparison of quiet and post-disturbance values on separate nights (Hilsenrath, pvt. communication).



In searching for physical mechanisms based on ozone to improve the forecastability of weather and climate, total ozone data bases are the wrong place to look. Rather, variations in ozone above the transport region should be investigated. Frederick (1977) has in fact theoretically modeled  $O_3$  response to solar UV changes over several solar rotations using the data of Heath (1973) and found 22% fluctuations about the norm in mesospheric ozone. Even the auroral effect produces approximately 7% change in mesospheric ozone. For the eleven year solar cycle both Callis and Nealy (1978) and Penner and Chang (1978) have calculated significant changes in stratospheric ozone, with the latter case showing agreement with Angell and Korshover. Although solar UV variability presents a major influence on natural ozone response, numerous calculations have also been published demonstrating the feasibility for ozone depletion by corpuscular bombardment (e.g. Frederick, 1976; Thorne, 1977; Heath et al., 1977).

Thus, there are apparent changes to stratospheric and mesospheric ozone induced by both transient and long-term solar variability. The question to be asked is if such ozone variability can affect tropospheric behavior. There are several new ideas linking ozone changes to stratospheric heating, alteration of the stratospheric circulation pattern, and eventual modification of the troposphere.

Fritz and Angell (1976) have suggested an auroral mechanism, whereby auroral IR heating may affect stratospheric temperatures by as much as  $2^\circ K$ . In support of this, one experiment showed stratospheric temperatures to be correlated with geomagnetic events (Ramakrishna and Seshamani, 1976). Others have looked at the effects of solar UV variability and corpuscular radiations. Stratospheric temperature variations due to changing solar UV flux intensities are substantial at heights above about 10 km, according to modeling results of Callis and Nealy (1978). At solar maximum, the greater heating of  $O_3$  by solar UV leads to calculated temperature increases up to a maximum of  $18^\circ K$  in the 40-55 km height range in three model cases and  $45^\circ K$  in the other two. Penner and Chang (1978), with a more comprehensive model, found the effects to be much smaller. Their calculated results are in rough agreement with observed stratospheric temperature variations over an 11-yr cycle (Schwentek, 1971; Zlotnik and Raswoda, 1976). No attempt was made in these calculations to determine how these temperature changes might affect atmospheric circulation.

Schoeberl and Strobel (1978) have considered circulation effects caused by corpuscular radiations, which deplete ozone to induce temperature changes on a regional scale. They conclude that even for the great PCA event of August 1972, the measured ozone depletion would be inadequate to cause significant change in the global mean circulation. However, Nastrom and Belmont (1978) have found zonal wind variations corresponding to the 27-day solar rotation period using cross-spectral analysis of 10.7 cm solar flux and rocketsonde data in the height range 25-65 km. They suggested that their results might be explained by Volland's (1977) land-sea heating model, which generates upward-propagating planetary waves; they pointed out that their latitudinal profile shows a wave structure near the stratopause characterized by amplitude maxima at  $25^\circ N$  and  $50^\circ N$  with approximately  $180^\circ$  phase lag, which is similar to the (2, -5) Rossby-Haurwitz wave given by Volland. Alternatively, Nastrom and Belmont say that direct ozone heating at stratospheric altitudes could produce the observed wind behavior.

Avery and Geller (1979) have theoretically investigated the effects of changing middle atmospheric winds on tropospheric planetary waves. The Avery and Geller calculations, based on a steady-state linear, quasi-geostrophic planetary wave model extending from the ground up to 100 km altitude with spherical geometry, indicate that the planetary wave amplitude can change significantly as its phase shifts east or west (i.e., ridges of highs and lows move E or W) mainly in response to jet stream amplitude changes. Here, variations in solar UV flux intensity over the 11-yr sunspot cycle and the 27-day solar rotation periods, and in solar corpuscular emissions, can alter the stratospheric ozone structure and produce changes in the temperature gradient through ozone heating. The temperature changes would then affect the strength of the polar night jet stream near the stratopause, causing amplitude variations and phase shifts in the planetary wave structure. Finally, variations in planetary wave structure would alter the position of principal storm tracks, with an attendant change in weather patterns.



Turning to the very long term Reid et al. (1976) have considered PCA's during Earth magnetic field reversals. During such periods (at intervals of the order  $10^6$  years) the Earth's field is sufficiently weak to permit solar proton bombardment globally following solar proton flares. They have considered how such bombardment on this more extensive scale might affect the biosphere, but also indicate possible influences on climate. Recently, Sonnenfeld (1978) has treated this problem on a more extensive scale, including all possible long-term modulations of solar UV impacting with the Earth's environment.

## 4.2 Middle Atmospheric Processes Involving Clouds and Aerosols

Clouds and aerosols may also respond to solar variability. Noctilucent clouds (NLC) are the highest in our atmosphere and occur near the summer polar mesopause ( $\sim 85$  km) in polar summer. Satellite data (Donahue et al., 1972) have shown NLC to be continuous over the polar cap ( $> 80^\circ$  latitude) for several weeks during polar summer. This is in contrast to ground based measurements, which have historically indicated temporally varying and spatially discontinuous structure, especially toward lower latitudes within the auroral zone. Little is known about the structure and composition of NLC. However, Hummel and Olivero (1976) have calculated radiative temperature changes at the Earth's polar surface on the basis of the satellite observations and conclude that changes of up to  $1^\circ\text{C}$  are possible, depending on particulate shape, size and concentration.

Statistical correlations of NLC with solar activity have been reported. For example, one observation claims that NLC rapidly dissipate following the onset of aurorae, and D'Angelo and Ungstrup (1976) postulated that this may be due to joule dissipation of large ionospheric electrical fields. Experimental evidence with a rocket-borne ion mass spectrometer (Goldberg and Witt, 1977) shows that metallic species of meteoric origin (especially  $\text{Fe}^+$ ) may be nucleation agents for NLC formation. Since the production of  $\text{Fe}^+$  ions is enhanced by precipitating particles, this may lead to increased NLC formation when other meteorological conditions are satisfied.

Within the stratosphere, nacreous (mother of pearl) clouds are occasionally observed, always near 30 km. They are most often sighted in northern latitudes during local winter, and can only be seen during the unique sunlight reflection conditions provided by sunrise and sunset. Meteorologists consider such clouds to be induced orographically, but this does not explain an absence of sightings at mid- and low latitudes. The impact of such clouds on coupling processes, though frequency of occurrence and spatial coverage, remains to be demonstrated, although effects originating through nucleation by solar proton events and other geomagnetic phenomena cannot be discounted at this time.

Solar and geomagnetic influences on cirrus clouds, highest in the troposphere, have also been postulated. These relatively frequent and well observed clouds can modulate the intensity of visible radiation reaching the lower atmosphere and surface. Roberts and Olsen (1973 a,b) have claimed that a cirrus cloud deck near 300 mb could cause local heating up to  $1^\circ\text{C}$  per day at high latitudes over a relatively warm ocean surface. The strong temperature gradients induced would then lead to significant circulation patterns within the troposphere. Roberts and Olsen also postulated that cirrus clouds are subject to solar activity effects through interaction with ionizing energetic corpuscular radiation, either directly or through secondary bremsstrahlung x-ray radiations. Johnson and Imhof (1975) have criticized this concept however, by arguing that the nearly constant ion pair production rate up to 28 km due to cosmic rays is rarely perturbed by x-ray radiations, even during major geophysical events.

Finally there is some evidence for submacroscopic aerosols within the middle atmosphere. These particulates have been deduced largely from the data of rocket-borne Gerdien and blunt probes. These data have exhibited huge order of magnitude changes of ion conductivity within the middle atmosphere as a function of latitude, time of day, season, and geomagnetic or solar disturbances.

Our lack of knowledge concerning such particulates is evidenced by the recently reported results of Mitchell et al. (1977). They conducted a series of four soundings to measure ion



conductivity through sunrise from White Sands, New Mexico (Figure 6). The large enhancements between 40 to 65 km are attributed to increases in ion mobility (reduction in ion size), since solar ionizing radiations have small effect below 65 km. This could be observed directly on the latter two flights in 1975, which used Gerdien probes capable of evaluating mobility from independent measurements of ion concentration and conductivity. The paper argues that the solar UV radiation must dissociate heavy ions or particulates to form smaller mass species during daytime, but a closer study of the profiles seem to indicate that the effect progressed from lower to higher altitudes, in conflict with that explanation. It is also not clear what effect the time separation ( $\sim 4$  years) between soundings may present.

More measurements are required to determine the character, distribution, size, and morphology of submacroscopic particulates within the middle atmosphere. If they exist in reasonable quantities, they may affect the transport and circulation properties of the middle atmosphere. For example, Hale (1977) has recently speculated a scavenging process whereby a few charged particulates of appropriate size, set in motion by electric field forces, could drag the neutral atmosphere and cause accumulation of trace constituents in localized regions. A knowledge of charged particulates and heavy ions is also important for the electrical considerations of the next section.

### 4.3 The Electrical Connection

In most classical texts on atmospheric electricity, e.g. Israel (1970, 1973), it is specified that the global electrical current is driven by global thunderstorm activity (about 1500 to 2000 storms at any given instance of time). In this classical picture, the basic circuit cannot be subject to perturbations from external sources, since the ionosphere acts as an "infinitely" conducting spherical shield. However, recent comparisons of the magnetospheric and atmospheric electric fields show them to be comparable magnitude (Sugiura, private communication, 1979). Since the ionosphere is not an infinite conductor in actuality, one might suspect penetration of the external magnetospheric fields into the middle and lower atmosphere, and vice versa. Furthermore, solar and magnetospheric fields connect in outer space, providing the possibility of a direct physical connection between solar active phenomena and the atmosphere.

The mapping of solar related phenomena to the atmospheric fair-weather electric circuit has been observed through correlative studies relating on the short term, to solar sector boundary passages and solar flares, and on the long-term to the 11-year sunspot cycle. The parameters studied are the return current (air-Earth current density,  $J_C$ ), the atmospheric vertical potential gradient (or electric field,  $E$ ), and the total ionospheric potential (potential difference between the ionosphere and ground,  $V_i$ ).

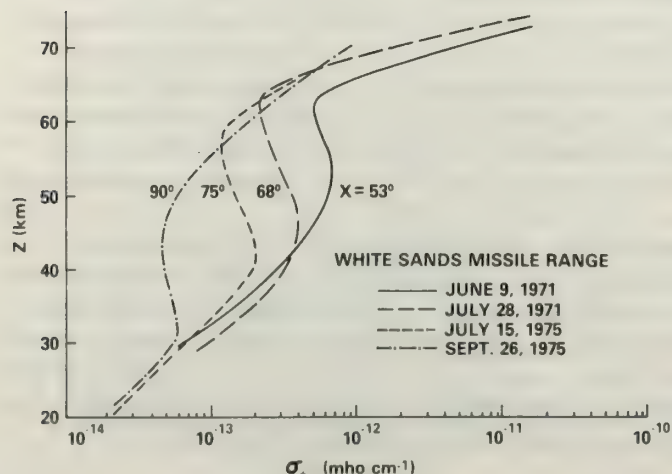


Figure 6. Positive ion conductivity ( $\sigma_+$ ) profiles with altitude for four solar zenith angles following sunrise. Rocket soundings occurred at White Sands Missile Range, New Mexico (after Mitchell et al. 1977).

Fair-weather measurements of  $E$  and  $J_C$  over an 11-year period at Zugspitze exhibit response to sector boundary crossings (Reiter, 1976; 1977). Reiter's superposed epoch analysis results show that both  $E$  and  $J_C$  increased by 20% or more on the two days following a  $-/+$  sector boundary crossing in maximum solar activity years, but on the same day as  $+/-$  crossings. A similar analysis of a 10-month period by Park (1976) indicates that the potential gradient at Vostok increased sharply by 20-30% beginning 3 days after boundary passages, and the response was similar for  $+/-$  and  $-/+$  boundary crossings. The difference in response time for the Vostok and Zugspitze electric fields remains unclear, although new modeling results by Roble and Hays (1979) may account for geographical differences in temporal lag.

For solar flares, ground measurements at Mauna Loa, Hawaii have shown that  $E$  and  $J_C$  increase significantly after solar flare eruptions. The potential gradient maximized 3-4 days after the flare day, while the air-earth current density reached maximum on +1 day; both quantities remained above normal for several days thereafter (Cobb, 1967). Reiter (1969, 1971, 1972) has shown that  $E$  and  $J_C$  measured at Zugspitze also increased following solar flare eruptions. Balloon measurements in the Arctic (Holzworth and Mozer, 1979) and the Antarctic (Cobb, 1978) have revealed that the potential gradient at 30-km decreases severalfold following solar flares. Such a reduction was anticipated by Herman and Goldberg (1976) on the basis of enhanced air conductivity due to solar proton ionization in the stratosphere.

Potential gradient enhancements have been observed to be correlated with solar radio noise bursts (associated with flares) at Zugspitze (Reiter, 1972), and at stations within the Arctic Circle (Sao, 1967). Märcz (1976) has found enhancements in  $E$  of 30-50% in Poland (approx.  $51^\circ\text{N}$ ) associated with geomagnetic storms (which generally follow solar flare eruptions), and Tanaka et al. (1977) saw similar enhancements in the Antarctic associated with auroral substorms.

Early long-term measurements of the fair-weather electric field suggested a stronger average field strength in years of maximum sunspot activity compared to minimum years at some European stations, but no apparent influence at others (c.f., the review by Herman and Goldberg, 1978b, p. 139). According to Mühleisen (1971), the total ionospheric potential measured over Germany for a complete solar cycle exhibits a positive correlation with annual mean sunspot number. To the extent that the total potential is proportional to global thunderstorm activity, one might therefore expect the latter quantity to be correlated with the 11-yr. sunspot cycle.

In contrast to the classical electrical picture that thunderstorms drive and control the fair weather electrical circuit, there is some emerging evidence that external influences may help modulate thunderstorms. At this time, the results are largely statistical and in many cases, questionable or controversial (Pittock, 1978; Herman and Goldberg, 1978b).

Markson (1971) was the first to observe solar sector boundary response and based his study on U.S. thunderstorm data collected during a solar minimum period covering approximately 52 sector boundary crossings. There seemed to be a definite preference for thunderstorms to occur from about 1 day before to one day after a  $+/-$  crossing, but no statistically significant response to  $-/+$  crossings. A more detailed analysis was made by Lethbridge (1979), who used a daily index derived from 30 years (1947-1976) of thunderstorm data from 102 weather stations in the U.S. In a superposed epoch analysis of three separate 10-yr sets of data, she found the strongest solar signal in the thunderstorm index for the winter months (Nov.-Mar.) in the latitude band  $40-45^\circ\text{N}$ , with peak activity occurring 1 day after  $+/-$  boundary crossings. With the index combined for all seasons and latitudes, and for both polarity crossings, Lethbridge could find no discernible response to sector boundaries.

Several different measures of thunderstorm activity have been utilized in studies of the response to solar flare eruptions, and all seem to show a positive response but with varying lag times. For example, Reiter (1969) used sferics counts in Germany (indicative of lightning flashes in thunderstorms within a 300-mile radius), and found a 57% increase in count rate peaking about 4 days after the eruption of solar flares with importance  $\geq 2$ . Other results include Bossolasco et al. (1973), who used the number of long-range radio direction-finding fixes recorded in the Mediterranean area as a measure of thunderstorm activity in that area for the 1961-1971 period, and showed a 50% increase in the daily number of fixes beginning one day and peaking 4 days after solar flare occurrences.



VLF whistler counts were used by Holzworth and Mozer (1979) as an indicator of high-latitude thunderstorm activity during the August 1972 solar-terrestrial events. They found a strong increase in count rate beginning about 12 hours after the class 3 flares of August 4 and August 7.

Finally, long-term studies of the variability of the annual number of thunderstorm days in different regions with the 11-yr solar cycle are replete with contradictions. The analysis of Siberian thunderstorms by Septer (1926) has been cited often as proof that middle-to-high latitude thunderstorm occurrence is directly correlated with annual mean sunspot number (correlation coefficient = +0.9). This result, however, has been severely criticized by Pittock (1978). Brooks' (1934) analysis from 57 years of England and Wales data suggests that annual mean sunspot numbers and annual number of thunderstorm days are uncorrelated, but Stringfellow (1974) found a strong positive correlation for the same region based on English thunderstorm occurrence in the years 1930-1973.

Several theories seeking an explanation for thunderstorm (lightning) response to solar activity have been proposed recently, but none bridge the final gap to explain how modifications in the electrical structure of the atmosphere can affect thunderstorm occurrence. Ney (1959) suggested that thunderstorm activity may be modulated by solar variability through alteration of the electrical state of the middle and lower atmosphere. Markson (1971, 1975, 1978), amplified this idea to evolve a theory which in its present state (1978) assumes that the electrical resistance of the atmosphere above thundercloud tops (the charging resistor, see Figure 7) will be lowered by enhanced ionization associated with incoming solar particles. The charging current is thereby increased leading to an enhanced ionospheric potential and fair-weather electric field which must adjust globally to the increased charging current.

D'Angelo (1978) has modified the global circuit by introducing a variable emf source representing the ionospheric potential at high latitudes. Fluctuations in this second driver force readjustments in the global circuit including the fair-weather electric field. He argues that the ionospheric potential is sensitive to solar wind electric fields, magnetospheric fields, etc., thereby introducing a coupling link within the solar-terrestrial system. These results are also consistent with the recent

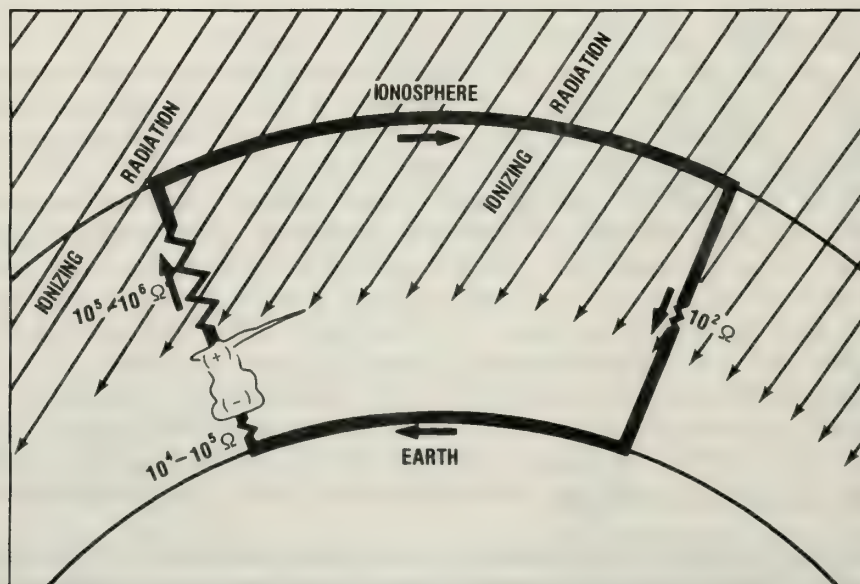


Figure 7. The atmospheric global electrical circuit. Large arrows indicate the flow of positive charge. Estimated resistances of circuit elements are given. The thunderstorm depicted represents the global electrical generator, i.e. the totality of all thunderstorms and sends current through the charging resistor ( $10^5 - 10^6 \Omega$ ) to the ionosphere. The cumulative effect of the global return current passes through the load ( $10^2 \Omega$ ) resistor) (after Markson, 1978).



analysis of Sugiura (1979, pvt. communication) which demonstrates the equivalent magnitudes of the fair-weather and magnetospheric electric fields, and help justify the concept of magnetospheric electrical field mapping into the stratosphere (Figure 8). New experimental results of Hale and Croskey (1979) appear to confirm D'Angelo's concept, that middle atmospheric electric fields in the auroral zone are sensitive to auroral phenomena. Hays and Roble (1979) have provided the most quantitative and sophisticated modeling effort to date. They have developed a global model for electrical parameters, which includes as input parameters orographic effects from the Earth's surface, the global thunderstorm distribution as observed from the DMSP Satellite, and the latitudinal distribution of cosmic ray flux. They, too, have calculated high latitude effects induced by solar active and/or magnetospheric phenomena, and have found significant perturbations on the global circuit properties.

Herman and Goldberg (1976, 1978a) have considered how cosmic rays and solar protons affect the local environment near thunderstorms, and if modifications in the local conductivity and electric fields can assist lightning generation. For the case of cosmic rays, the changes appear quantitatively reasonable based on the school of thought (c.f., Chalmers, 1967) that an increase in the fair-weather field enhances the probability of thunderstorm formation under appropriate meteorological conditions. Thus, solar-controlled variation in cosmic ray intensity, especially over the 11-yr. cycle, may modulate the fair-weather field and hence the rate of thunderstorm occurrence. The more difficult question concerns how solar protons, which typically are absorbed above 20-25 km, could affect tropospheric electrical structure. Figure 9 compares penetration depths for protons during several large PCA events of the last decade, and compares the effect of cosmic rays over a solar cycle.

Follin et al. (1977) have taken a different approach to explain lightning generation in thunderstorms. Beginning with the premise of a pre-existing thundercloud, they argue that meteorological dynamics cannot generate sufficient electric potential to equal or exceed the breakdown voltage of the atmosphere. If however, the cascading muons from a high energy cosmic ray pass through the thundercloud they can provide an ionized path for a lightning bolt to follow. If there is any solar control of the occurrence rate of cosmic ray air showers, the incidence of lightning should correlate with solar variables.

The final scenario, what happens to the meteorology by changes in atmospheric electrification, is open to speculation. Here, various suggestions have postulated improved conditions for hail formation, water droplet growth and enhanced rain, etc. These concepts await physical validation.

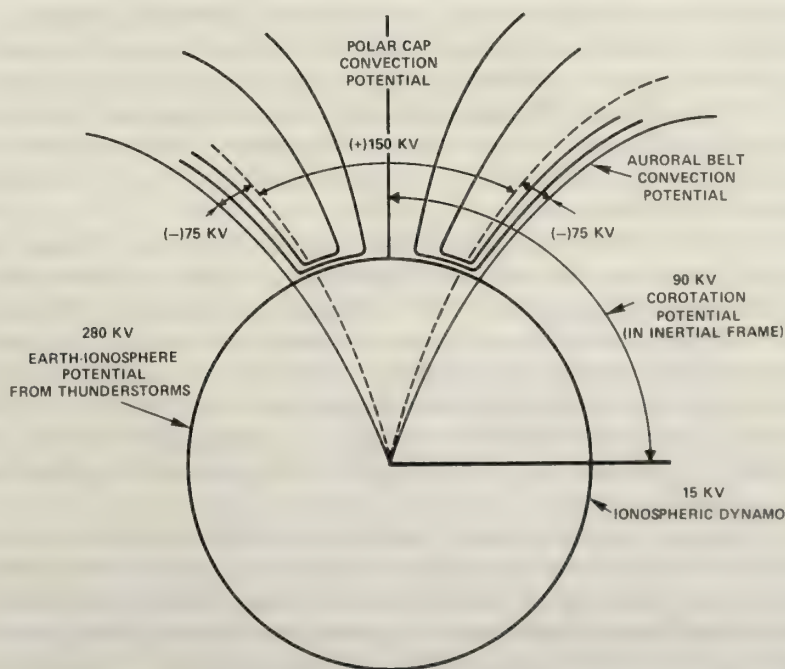


Figure 8. Schematic illustrating sources and magnitudes of electric fields in near Earth space. (Sugiura, pvt. communication).

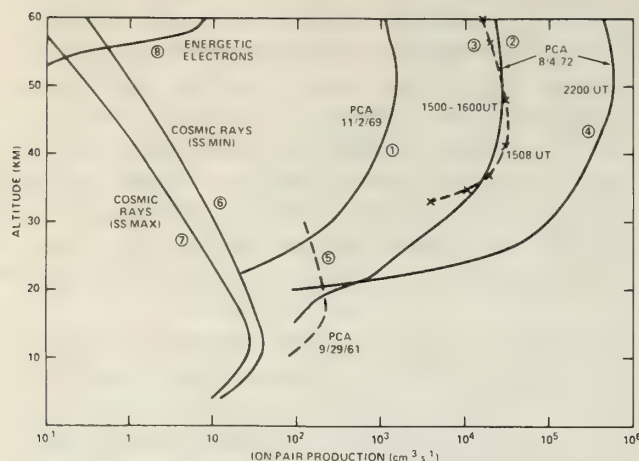


Figure 9. Ion pair production rates due to solar protons (curves 1-5) galactic cosmic rays at sunspot minimum (curve 6) and sunspot maximum (curve 7), and energetic electrons (curve 8). These curves are based on calculations from various sources as given in Herman and Goldberg (1978a).

## 5. INDICATORS POTENTIALLY USEFUL FOR FORECASTING

For sun-weather relationships to be useful in forecasting, there must be identifiable (observable) solar activity precursors or indicators that herald changes in weather and climate parameters.

The parameters of interest to the short-term forecast user include air temperature, winds, cloud cover and rainfall including thundershowers, supplemented by additional warnings and updates for severe storm conditions such as blizzards, hail and thunderstorms, tornadoes and hurricanes. Prediction of long-term climatic trends center principally on precipitation amounts (including droughts), temperatures and frequencies of severe storms.

Solar activity indicators potentially useful as precursors include sunspots; solar flare eruptions; enhancements of short (X-rays, EUV, UV) and long (solar radio noise) wavelength electromagnetic radiations; energetic particle emissions; geomagnetic storm commencements; solar magnetic sector boundary crossings; coronal holes (M regions); ground level enhancements (GLE) and Forbush decreases in cosmic ray intensity; and possibly active plage regions. Quasi-periodicities in solar activity (e.g., the 11- and 22-yr sunspot cycles; 27-day solar rotation rate) may prove to be useful predictor tools when a greater understanding of sun-weather physics is available.

Practically all of these have been utilized at one time or another in correlation studies with weather and climate parameters. Some apparent successes have been achieved, but ignorance of the coupling processes involved plus failures to verify the results with independent data sets place the successes in doubt.

Sunspot numbers, expressed as monthly or annual averages and sometimes as smoothed or running weighted means over several years, have been correlated with essentially every weather and climate parameter of interest (c.f., Herman and Goldberg, 1978b). Correspondingly, these parameters averaged over the same time spans have been used for single-station analyses, averaged over the globe or over specific geographic regions. The most promising finding to date appears to be a relationship between the 22-yr sunspot cycle and droughts in the western United States (Mitchell et al., 1977). This result seems to be sufficiently unassailable that the next major U.S. drought period should occur in the late 1990's, but without a physical explanation even this prediction must be made with caution.

The potential of sunspot observations for short term weather forecasting derives from their utility in prediction of ionospheric and geomagnetic storms. It is well known that a large active sunspot group appearing on the sun's east limb may (within the following two weeks) spawn one or more large flares with their accompanying outbursts of energy in the form of X-rays, UV, and energetic particles. The consequent effects on the Earth's ionosphere and magnetic field are also well



known. If it can be shown that the energetic outbursts from the sun affect weather parameters either directly or indirectly, sunspot observations would find a place in weather forecasting.

However, sunspots give only the first clue. More detailed observations of activity suspected to be associated with atmospheric processes can lead to a better connection. For example, Blackshear and Tolson (1978) have shown recently that monthly averaged global total atmospheric ozone varied with the average solar Ly  $\alpha$  flux intensity with a correlation coefficient of 0.94 (based on ten months of data), but with monthly averaged sunspot number the coefficient was only 0.71. In this case, however, the coefficient of correlation with 10.7 cm radio flux was even less, at 0.66. If changes in global ozone concentration affect the weather through modulation of the heat flux to the lower atmosphere or stratospheric circulation in an understandable way, monitoring of solar Ly  $\alpha$  could become a useful indicator.

Coronal holes may be of interest in this regard, since they are the visible images of M regions made in the EUV and x-ray portions of the spectrum. They are the source of high-speed corpuscular radiation as well as short wave electromagnetic radiation, and they repeat with a 27-day recurrence because a hole may persist for up to 8 or 10 solar rotations. Ionospheric physicists have been using the M-region concept for years to predict recurrent geomagnetic and ionospheric storms (and radio absorption events). Again, if the enhanced radiation emissions can be demonstrated to affect weather, coronal hole observations could be a factor in forecasting. Forecasts could conceivably be made on the basis of the 27-day recurrence, but when coronal hole observations indicated that the solar feature responsible for the 27-day pattern had disappeared, the persistence forecast would have to be modified or abandoned. Musman and Altcock (1978) have pointed out that an improvement over a strictly 27-day recurrence might be had by taking into account the brightness of the Fe XIV line (5303 Å) associated with coronal holes.

Solar flare eruptions appear to be followed by several atmospheric responses with lag times ranging from several hours to a few days. Included are significant atmospheric pressure enhancements and temperature decreases near the 300-mb level (Schuurmans and Oort, 1969); increased thunderstorm activity in the Mediterranean area (Bossolasco et al., 1973); potential gradient enhancements in Bavaria (Reiter, 1969; 1971); increased total ionospheric electric potential (Reiter, 1972; Mühleisen and Reiter, 1973); decreased ozone content at 4 mb (Heath et al., 1977); decreased seasonal average total ozone (Angell and Korshover, 1975); redistribution of surface pressure patterns (Duell and Duell, 1948; Schuurmans and Oort, 1969); and changes in the 500-mb height (Schuurmans, 1965). Additional study of solar flare effects, and an extension of the results to changes in weather parameters of interest are required before solar flare observations can be utilized for improving weather forecasts. Consideration should be given to this possibility, since the global network of solar observatories provide continuous monitoring of the sun.

Solar magnetic sector boundaries sweeping past the Earth are apparently accompanied or followed by changes in atmospheric vorticity (Wilcox et al., 1973; Larsen and Kelley, 1977); the electric potential gradient in the Antarctic (Park, 1976) and in Bavaria (Reiter, 1977); and the total ionospheric electric potential (Reiter, 1977). Though sector boundary passages offer a potential for predicting changes in the vorticity area index (VAI), Williams (1978) has pointed out two disadvantages for meteorological forecasting. First, the VAI is not routinely used in the atmospheric sciences and there is no background information on its typical behavior. Second, according to Williams, the VAI is a single isolated parameter which cannot provide information about what is happening to other atmospheric parameters while the VAI itself is responding to solar modulation. Even without this complication, some researchers have failed to find an association between sector boundaries and the VAI (Williams, 1978; Bhatnagar and Jakobsson, 1978). This failure, however, may be due to the method of analysis used (Larsen, 1978). Additional study is therefore required before sector boundaries can be utilized for forecast purposes. This solar parameter is a potential candidate because it would provide several days of lead time. That is, atmospheric responses occur typically one or two days after a boundary sweeps past the Earth, but the solar-rooted foot of the boundary crosses the central meridian of the sun  $4\frac{1}{2}$  days prior to that passage.

An interesting possibility for long-term climate forecasting (at least regarding colder than normal temperatures) has recently emerged. It has been shown by Eddy et al. (1977) that the rotation



of the sun accelerated between 1625-26 and 1642-44, just prior to the Maunder minimum in sunspot activity associated with the coldest part of the Little Ice Age. The differential rotation was most dramatically affected because the acceleration was observed only for spots within  $15^\circ$  of the solar equator. Recent analysis of Thomas Harriott's sunspot data by Herr (1978) indicates that the rotation rate was slower in 1611-13 than it was in 1625-26. This implies that the acceleration was even greater than Eddy et al. had found for the decades prior to the Maunder minimum.

With apologies to the solar physicists, it may be that the solar energy previously vented by the sun in association with sunspots was now being used for accelerating its rotation rate. In such case, there would be fewer or no sunspots, and less energy variations available to affect the Earth's atmosphere in whatever unknown ways it might otherwise do so. For long-term forecasting of climate variations, then, the rotation rate of the sun becomes a candidate. Additional investigation of recent variations in this parameter would therefore be of interest.

## 6. SUMMARY AND CONCLUSIONS

We have briefly considered the various radiations reaching the middle atmosphere in response to solar activity, and their known effects on neutral and ionic species of that region. Those responses have then been evaluated as possible coupling mechanisms connecting the upper and lower atmosphere. Two general categories emerge. The first is the role of the stratosphere and mesosphere as a buffer region, wherein small perturbations of local parameters can be amplified into major effects affecting the transport or conversion of energy from external sources reaching the region. These effects involve ozone, aerosols, and to a lesser extent, charged species. The major effects involve heating, circulation patterns, and possible long term effects on climate in both the stratosphere and troposphere.

The second, the electrical connection, involves changes in the global and local electrical structure of the atmosphere induced by solar activity. These effects are more direct and may possibly bypass the stratosphere altogether. The responses in the atmospheric circuit to local changes are nearly instantaneous, and therefore are looked on with more promise for possible influences on weather systems. Newly emerging experimental evidence indicates that atmospheric electric fields, especially at high latitudes and in the upper atmosphere, are responsive to solar and geomagnetic phenomena. The corpuscular radiations previously discussed are known to strongly enhance the local atmospheric electrical conductivity and ionizations at stratospheric and mesospheric altitudes. However, the electric field perturbations are not easily explained by conventional atmospheric electricity considerations.

Several solar activity indicators have been identified as potential candidates for predicting weather or climate changes. These include sunspots, solar flare eruptions, coronal holes, and solar magnetic sector boundary passages for short term forecasts, and sunspots, solar rotation rate, and possibly solar constant variations for long term climate predictions. Unfortunately, it appears that the present state of sun-weather art precludes immediate application of these potential candidates.

The various ideas presented here are primarily hypothetical and are not yet well enough founded to apply to forecasting. Experiments and additional studies are needed to investigate these and other more promising concepts and establish the importance of each conceived idea. Only then will it be possible to develop methods to incorporate sun-weather relationships into prediction programs.

## REFERENCES

- Angell, J. L. and J. Korshover (1973): Quasi-biennial and long-term fluctuations in total ozone, Monthly Weath. Rev., 101:426.
- Angell, J. L. and J. Korshover (1975): Global analysis of recent total ozone fluctuations. Rept. of NOAA Air Resources Laboratory (Reviewed in EOS, 56:584).
- Arendt, P. R. and E. M. Frisby (1968): Possible relation of a specific ionospheric event to simultaneous meteorological data, Nature, 219:475.
- Avery, S. K. and M. A. Geller (1979): Effects of middle atmosphere winds on planetary waves in the troposphere, submitted to Geophys. Res. Let.
- Baker, D. M. and K. Davies (1969): F2-region acoustic waves from severe weather, J. Atmos. Terr. Phys., 31:1345.
- Bauer, S. J. (1957): A possible tropospheric-ionospheric relationship, J. Geophys. Res., 62:425.
- Bauer, S. J. (1958): An apparent ionospheric response to the passage of hurricanes, J. Geophys. Res., 63:265.
- Bering, E. A., J. R. Benbrook, D. Detrick, T. J. Rosenberg, M. J. Rycoft, M. A. Saunders and W. R. Sheldon (1979): Electric fields, energetic electron precipitation and VLF radiation during a simultaneous magnetospheric substorm and atmospheric thunderstorm, submitted for publication.
- Bhatnagar, V. P. and T. Jakobsson (1978): Lack of effects of solar magnetic sector crossings on the troposphere, Geophys. Res. Let., 5:180.
- Blackshear, W. T. and R. H. Tolson (1978): High correlations between variations in monthly averages solar activity and total atmospheric ozone, Geophys. Res. Let., 5:921.
- Boeck, W. L. (1978): Meteorological consequence of atmospheric krypton-85, Science, 200:727.
- Bossolasco, M., I. Dagnino, A. Elena and G. Flocchini (1973): The thunderstorm activity over the Mediterranean area, Revista Italiana di Geofisica, 12:21.
- Brooks, C. E. P. (1934): The variation of the annual frequency of thunderstorms in relation to sunspots, Quart. J. R. Meteorol. Soc., 60:153.
- Callis, L. B. and J. E. Nealy (1978): Solar UV variability and its effect on stratospheric thermal structure and trace constituents, Geophys. Res. Let., 5:249.
- Chalmers, J. A. (1967): Atmospheric Electricity, 2nd edition, Pergamon Press, Oxford.
- Chameides, W. L., D. H. Stedman, R. R. Dickerson, D. J. Busch, and R. J. Cicerone (1977): NO<sub>x</sub> production in lightning, J. Atmos. Sci., 34:143.
- Cobb, W. E. (1967): Evidence of a solar influence on the atmospheric electric elements at Mauna Loa Observatory, Monthly Weather Rev., 95:12.
- Cobb, W. E. (1978): Balloon measurements of the air-earth current density at the South Pole before and after a solar flare, preprint from AMS Conf. on Cloud Physics and Atmospheric Electricity, Issaquack, Washington.
- D'Angelo, N. (1978): Thunderstorms, ionosphere to ground potential drop, and the solar wind sector structure, U. of Iowa Report, 78-24.
- D'Angelo, N. and E. Ungstrup (1976): On the occurrence of widely observed noctilucent clouds, J. Geophys. Res., 81:1777.
- Davies, K. (1965): Ionospheric Radio Propagation, National Bureau of Standards Monograph #80.
- Donahue, T. M., B. Guenther, and J. E. Blamont (1972): Noctilucent clouds in daytime; circum-polar particulate layers near the summer mesopause, J. Atmos. Sci., 29:1205.
- Dubin, M. and E. P. Zipf (1977): Nitrogen fixation from atmospheric discharges as the prime source of atmospheric nitrous oxide and soil nitrates, EOS, 58:464.
- Duell, B. and G. Duell (1948): The behavior of atmospheric pressure during and after solar particle invasions and solar ultraviolet invasions, Smithsonian Misc. Collection 110, 34 pp.



- Eddy, J. A., P. A. Gilman and D. E. Trotter (1977): Anomalous solar rotation in the early 17th century, Science, 198:824.
- Follin, J. W., Jr., E. P. Grey and K. Yu (1977): The connection between cosmic rays and thunderstorms, EOS, 58:1220.
- Frederick, J. E. (1976): Solar corpuscular emission and neutral chemistry in the Earth's middle atmosphere, J. Geophys. Res., 81:3179.
- Frederick, J. E. (1977): Chemical response of the middle atmosphere to changes in the ultraviolet solar flux, Plant. Space Sci., 25:1.
- Fritz, S. and J. K. Angell (1976): Temperature and wind variation in the tropical upper stratosphere and lower mesosphere during a sunspot cycle, J. Geophys. Res., 81:1051.
- Georges, T. M. (1968): HF doppler studies of travelling ionospheric disturbances, J. Atmos. Terr. Phys., 30:725.
- Gherzi, E. (1950): Ionosphere and weather, Nature, 165:38.
- Goldberg, R. A. (1979): An experimental search for causal mechanisms in sun/weather-climatic relationships in Proceedings of the Symposium/Workshop on Solar Terrestrial Influences on Weather and Climate, Columbus, Ohio, July 1978, ed. by B. M. McCormac and T. Seliga.
- Goldberg, R. A. and G. Witt (1977): Ion composition in a noctilucent cloud region, J. Geophys. Res., 82:2619.
- Hale, L. C. (1977): Particulate transport through the mesosphere and stratosphere, Nature, 268:710.
- Hale, L. C. and C. L. Croskey (1979): An auroral effect on the fair weather electric field, Nature, 278:239.
- Heath, D. F. (1973): Space observations of the variability of solar irradiance in the near and far ultraviolet, J. Geophys. Res., 78:2779.
- Heath, D. F., A. J. Krueger, and P. J. Crutzen (1977): Solar proton event influence on stratospheric ozone, Science 197:886.
- Heath, D. F. and M. P. Thekaekara (1977): Measures of the solar spectral irradiance between 1200 and 3000A, in The Solar Output and its Variations, ed. by O. R. White, Colorado Associated University Press, Boulder, Colorado.
- Herman, J. R. and R. A. Goldberg (1976): Solar activity and thunderstorm occurrence, EOS, 57:971.
- Herman, J. R. and R. A. Goldberg (1978a): Initiation of non-tropical thunderstorms by solar activity, J. Atmos. Terr. Phys. 40:121.
- Herman, J. R. and R. A. Goldberg (1978b): Sun, Weather and Climate, NASA SP426.
- Herr, R. B. (1978): Solar rotation determined from Thomas Hariott's sunspot observations of 1611 to 1613, Science 202:1079.
- Hines, C. O. and I. Halevy (1977): On the reality and nature of a certain sun-weather correlation, J. Atmos. Sci., 34:382.
- Holzworth, R. H. and F. S. Mozer (1979): Direct evidence of solar flare modification of stratospheric electric fields, J. Geophys. Res., 84:363.
- Hummel, J. R. and J. J. Olivero (1976): Satellite observation of the mesospheric scattering layer and implied climatic consequences, J. Geophys. Res., 81:3177.
- Hung, R. J., G. L. Rao, R. E. Smith, G. S. West and B. B. Hensen (1975): Ionospheric disturbances during severe weather activity, in Proc. Symp. Effects of the Ionosphere on Space Systems and Communications, Naval Research Laboratory.
- Israel, H. (1970): Atmospheric Electricity, Volume 1: Fundamentals, Conductivity, Ions, Israel Program for Scientific Translations, Jerusalem NTIS Doc. TT-67-51394/1.
- Israel, H. (1973): Atmospheric Electricity, Volume 2: Fields, Charges, Currents, Israel Program for Scientific Translations, Jerusalem, NTIS Doc. TT-67-51394/2.
- Johnson, R. G. and W. L. Imhoff (1975): Direct satellite observations on bremsstrahlung radiation as a technique to investigate its role in meteorological processes, NASA SP366, ed. by W. R. Bandeen and S. P. Maran.



- King, J. W. (1975): Sun-weather relationships, Astronautics and Aeronautics, p. 10, April.
- Larsen, M. F. (1978): Comments on the paper "Lack of effects of solar magnetic sector crossings on the troposphere" by V. P. Bhatnagar and T. Jakobsson, Geophys. Res. Let., 5:619.
- Larsen, M. F. and M. C. Kelley (1977): A study of an observed and forecasted meteorological index and its relation to the interplanetary magnetic field, Geophys. Res. Let., 4:337.
- Lethbridge, M. (1979): Thunderstorm frequency and solar sector boundaries in Proceedings of Symposium/Workshop on Solar Terrestrial Influences on Weather and Climate, July 1978, Columbus, Ohio, ed. by B. M. McCormac and T. Selega.
- Märcz, F. (1976): Links between atmospheric electricity and ionospheric absorption due to extra-terrestrial influences, J. Geophys. Res., 81:4566.
- Markson, R. (1971): Considerations regarding solar and lunar modulation of geophysical parameters, atmospheric electricity, and thunderstorms, Pure Appl. Geophys., 84:161.
- Markson, R. (1975): Solar modulation of atmospheric electrification through variation of the conductivity over thunderstorms, in NASA SP-366, ed. by W. R. Bandeen and S. P. Maran.
- Markson, R. (1978): Solar modulation of atmospheric electrification and possible implications for the sun-weather relationship, Nature, 273:103.
- Meadows, A. J. (1975): A hundred years of controversy over sunspots and weather, Nature, 256:95.
- Mitchell, J. D., R. S. Sagar and R. O. Olsen (1977): Positive ions in the middle atmosphere during sunrise conditions, Space Res., 17:199, Pergamon Press, Oxford.
- Mitchell, J. M. Jr., C. W. Stockton, and D. M. Meko (1977): Drought cycles in the United States and their relation to sunspot cycle since 1700 A.D., EOS, 58:694.
- Mühleisen, R. (1971): Neue Ergebnisse und Probleme in der Luftelektrizität, Zs. Geophysik, 37:759.
- Mühleisen, R., and R. Reiter (1973): Atmospheric electric data during August 4th-12th, 1972, in Collected Data Reports on August 1972 Solar-Terrestrial Events Rpt. UAG-28, Part III, World Data Center A, Boulder, Colorado, p. 809.
- Musman, S. and R. C. Altcock (1978): Recurrent geomagnetic disturbances and coronal holes as observed in Fe XIV 5303A, J. Geophys. Res., 83:4817.
- Nastrom, G. D. and A. D. Belmont (1978): Preliminary results on 27-day solar rotation variation in stratospheric zonal winds, Geophys. Res. Let., 5:665.
- Ney, E. P. (1959): Cosmic radiation and the weather, Nature 183:451.
- Noxon, J. F. (1976): Atmospheric nitrogen fixation by lightning, Geophys. Res. Let. 3:463.
- Park, C. G. (1976): Solar magnetic sector effects on the vertical atmospheric electric field at Vostok, Antarctica, Geophys. Res. Let., 3:475.
- Park, C. G. and R. A. Helliwell (1978): Magnetospheric effects of power line radiation, Science, 200:727.
- Penner, J. E. and J. S. Chang (1978): Possible variations in atmospheric ozone related to the eleven year solar cycle, Geophys. Res. Let., 5:817.
- Pittock, A. B. (1978): A critical look at long term sun-weather relationships, Rev. of Geophys. & Space Sci., 16:400.
- Ramakrishna, S. and R. Seshamani (1976): Day night dependence of geomagnetic activity effects on mesospheric temperature, J. Geophys. Res., 81:6173.
- Reagan, J. B. (1977): Ionization processes in Dynamical and Chemical Coupling of the Neutral and Ionized Atmosphere, ed. by B. Grandal and J. A. Holtet, p. 145, D. Reidel Publishing Co., Dordrecht, Holland.
- Reid, G. C., T. E. Holzer, I. S. A. Isaaksen and P. J. Crutzen (1976): The influence of ancient solar proton events on the evolution of life, Nature, 259:177.
- Reiter, R. (1969): Solar flares and their impact on potential gradient and air-earth current characteristics at high mountain stations, Pure Appl. Geophys., 72:259.
- Reiter, R. (1971): Further evidence for impact of solar flares on potential gradient and air-earth current characteristics at high mountain stations, Pure Appl. Geophys., 86:142.

- Reiter, R. (1972): Case study concerning the impact of solar activity upon potential gradient and air-earth current in the lower troposphere, Pure Appl. Geophys., 94:218.
- Reiter, R. (1976): The electrical potential of the ionosphere as controlled by the solar magnetic sector structure, Die Naturwissenschaften, 63, Part 4: 192.
- Reiter, R. (1977): The electrical potential of the ionosphere as controlled by the solar magnetic sector structure, result of a study over the period of a solar cycle, J. Atmos. Terr. Phys. 39:95.
- Roberts, W. O., and R. H. Olson (1973a): New evidence for effects of variable solar corpuscular emission on weather, Rev. Geophys. Space Phys., 11:731.
- Roberts, W. O. and R. H. Olson (1973b): Geomagnetic Storms and wintertime 300-mb trough development in the North Pacific - North America Area, J. Atmos. Sci., 30:135.
- Roble, R.G. and P.E. Hays (1979): Electrical coupling between the upper and lower atmosphere, in Proceedings of the Symposium/Workshop on Solar Terrestrial Influences on Weather and Climate, July, 1978, Columbus, Ohio, ed. by B. McCormac and T. Selega.
- Sao, K. (1967): Correlations between solar activity and the atmospheric potential gradient at the Earth's surface in the polar regions, J. Atmos. Terr. Phys., 29:213.
- Schoeberl, M. R. and D. F. Strobel (1978): The response of the zonally averaged circulation to stratospheric ozone reductions, J. Atmos. Sci., 35:1951.
- Schuermans, C. J. E. (1965): Influence of solar flare particles on the general circulations of the atmosphere, Nature, 205:167.
- Schuermans, C. J. E. and A. H. Oort (1969): A statistical study of pressure changes in the troposphere and lower stratosphere after strong solar flares, Pure Appl. Geophys., 75:233.
- Schwentek, H. (1971): The sunspot cycle 1958/70 in ionospheric absorption and stratospheric temperature, J. Atmos. Terr. Phys., 33:1839.
- Sechrist, C. (1979): Workshop Report on the role of minor atmospheric constituents in solar-terrestrial influences on weather and climate in Proceedings of Symposium/Workshop, Solar Terrestrial Influences on Weather and Climate, Columbus, Ohio, July 1978, ed. by B. M. McCormac and T. Selega.
- Septer, E. (1926): Sonnenflecken und Gewitter in Siberien, Meteorol. Z. 43:229.
- Sonnenfeld, T. (1978): Effects of a variable sun at the beginning of the Cenozoic era, Climate Change, 1:355.
- Stringfellow, M. F. (1974): Lightning incidence in Britain and the solar cycle, Nature, 249:332.
- Swider, W. and T. Keneshea (1973): Decrease of ozone and atomic oxygen in the lower mesosphere during a PCA event, Planet. Space Sci., 21:1969.
- Tanaka, Y., T. Ogawa and M. Kodama (1977): Stratospheric electric fields and currents measured at Syowa Station, Antarctica - 1. The vertical component, J. Atmos. Terr. Phys., 39:523.
- Thorne, R. M. (1977): Energetic radiation belt electron precipitation: a natural depletion mechanism for stratospheric ozone, Science, 195:287.
- Volland, H. (1977): Periodic variations of the solar radiation - a possible source of solar activity-weather effects, J. Atmos. Terr. Phys., 39:69.
- Weeks, L. H., R. S. Cuikay and J. R. Corbin (1972): Ozone measurements in the mesosphere during the solar proton event of 2 November 1969, J. Atmos. Sci., 29:1138.
- Wilcox, J. M. (1975): Solar activity and the weather, J. Atmos. Terr. Phys., 37:237.
- Wilcox, J. M., P. H. Scherrer, L. Svalgaard, W. O. Roberts and R. H. Olson (1973): Solar magnetic structure: Influence on stratospheric circulation, Science, 180:185.
- Williams, R. G. (1978): A study of the energetics of a particular sun-weather relation, Geophys. Res. Lett., 5:519.
- Zlotnik, B. and W. Rozwoda (1976): The influence of solar activity on the temperature of the stratosphere, Artificial Satellites, 2:55.



# THE IMPACT OF SUN-WEATHER RESEARCH ON FORECASTING

Miguel F. Larsen  
Division of Atmospheric Sciences  
&  
School of Electrical Engineering  
Cornell University  
Ithaca NY 14853

We examine the possible impact of sun-weather research on forecasting. The type of knowledge of the effect that we now have is evaluated to determine if it is in a form that can be used for forecasting purposes. We conclude that our present understanding of the effect does not lend itself readily to applications for forecast purposes. The limits of present predictive skill are examined and it is found that skill is most lacking for prediction of the smallest scales of atmospheric motion. However, it is not expected that sun-weather research will have any significant impact on forecasting the smaller scales since predictability at these scales is limited by the finite grid size resolution and the time scales of turbulent diffusion. The predictability limits for the largest scales are on the order of several weeks although presently only a one week forecast is achievable. This would be the most fruitful scale for the application of the research results.

## 1. INTRODUCTION

In assessing the possible impact of sun-weather research on forecasting, we need to make a serious evaluation of the results that we have as well as those that will be needed in the future if the results are to have any value in the forecasting process. Up til the present, the interest of the researchers in the field has been scientific rather than practical. This is not to say that the two cannot be made compatible, but some effort will be needed on the part of both researchers and forecasters to accomplish this. The fact that a very small solar signal can create a "noticeable" effect at tropospheric heights is a very interesting phenomenon from the point of view of those studying the atmosphere, but from the point of view of the forecaster, this will only be interesting if knowledge of this effect will increase his daily forecast skill score. The reader should notice that we have used the word "noticeable" in describing the sun-weather effect. In the literature, the effect is most often described as significant, but this refers to being statistically significant rather than significant for forecast purposes. We have to ask whether the effect is also significant enough to affect the routine forecast products. This paper is organized as follows.



We will first review what we know about the sun-weather effect as it relates to possible inputs for forecasting procedures; second, we will try to define the exact nature of the input which would be needed for this to be of value; and last, we will estimate the limits of forecast improvement which could be expected if these procedures are implemented.

## 2. PROCESS OF FORECAST PRODUCTION

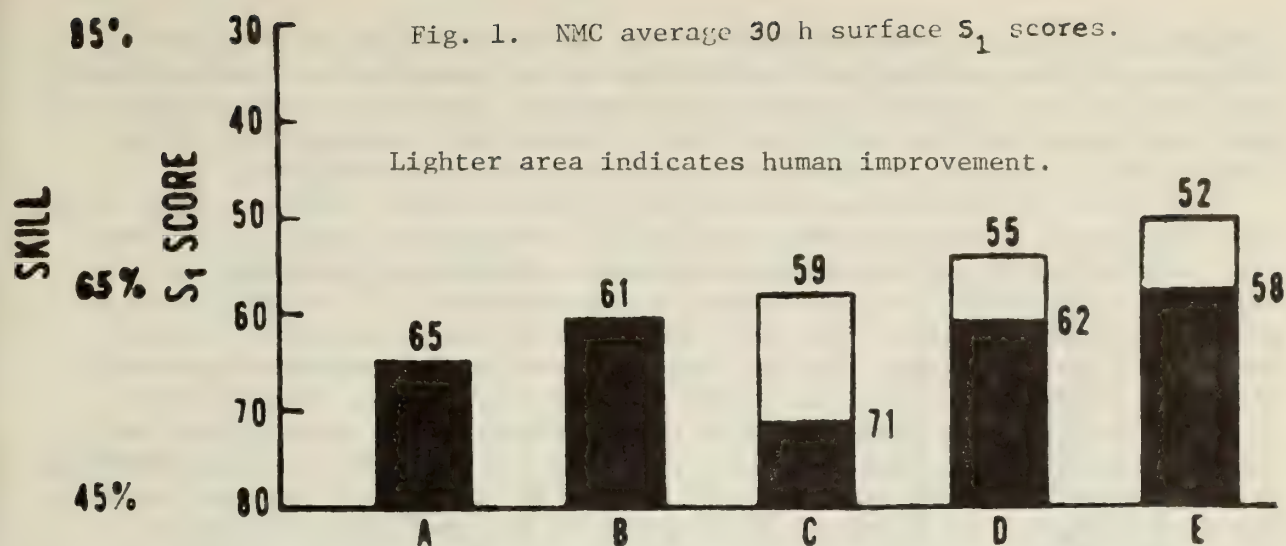
The process of producing a forecast consists of several stages (Bengtsson, 1976). First, data is taken at surface stations once every three hours and by means of radiosonde once every 12 hours. Satellite data and pilot reports are also available at various times. This data is analyzed and distributed by facsimile and teletype to the various forecast centers. The radiosonde data is used as an initial data field which is the input for various numerical models the most accurate of which is the 6 level primitive equation model. The output from the models is conveyed to the forecast centers by means of facsimile and teletype. These model outputs which include the predicted surface pressure, 500 mb. geopotential heights, vertical velocity field, predicted precipitation pattern, and 700 mb. height field for a 12, 24, 36, and 48 hour forecast are then used as guidance along with the synoptic chart analyses. At that point the individual forecaster decides which parameters are important for the forecast users and assimilates all his data into a final prediction.

In evaluating the possible impact of sun-weather research on this process, we must somehow try to evaluate whether or not knowledge of the effect of solar variability on the atmosphere will make a significant difference in relation to the skill of that final prediction. This will depend on various factors such as the parameter which is being forecast. In some cases it would be more important to know whether or not there will be precipitation in the forecast period. In other cases, the maximum and minimum temperatures or possibly the mean temperature will be the critical factor. This will depend on who the user is. These questions are of a complex nature, and in order to deal with them we will restrict our attention for the most part to the possible impact of sun-weather research on the skill of the forecast product produced by the numerical models. This is justified by the fact that the forecaster can usually improve the skill of the forecast by about 5 points on the  $S_1$  scale, where

$$S_1 = \frac{\sum |E_g|}{\sum |G_1|} \times 100 \quad (1)$$

$E_g$  is the error in the observed pressure gradient and  $G_1$  is the maximum of the observed and forecast pressure gradients. The 5 pt. forecast improvement may be attributable to the forecaster's knowledge of the local climatology, or orographic effects or possibly his experience of consistent errors made by the model in certain situations. It is not likely that this improvement is due to the forecaster's awareness of sun-weather effects since few meteorologists have access to solar data on a routine basis. This could possibly be worth investigating in the future. Figures 1 and 2 taken from Cooley and Derouin (1972) indicate the increase in forecast skill achieved with the improvement of the numerical forecast models. The hatched area is the increase in skill due to human improvement of the forecast.

Fig. 1. NMC average 30 h surface  $S_1$  scores.



- A. Pre-NWP.
- B. April 1956 to February 1964. Based on NWP 500-MB barotropic and 3-level baroclinic guidance.
- C. March 1964 to May 1966. Based on Reed surface guidance.
- D. June 1966 to August 1967. Based on P.E. surface guidance.
- E. September 1967 to December 1971. Based on P.E. surface guidance.

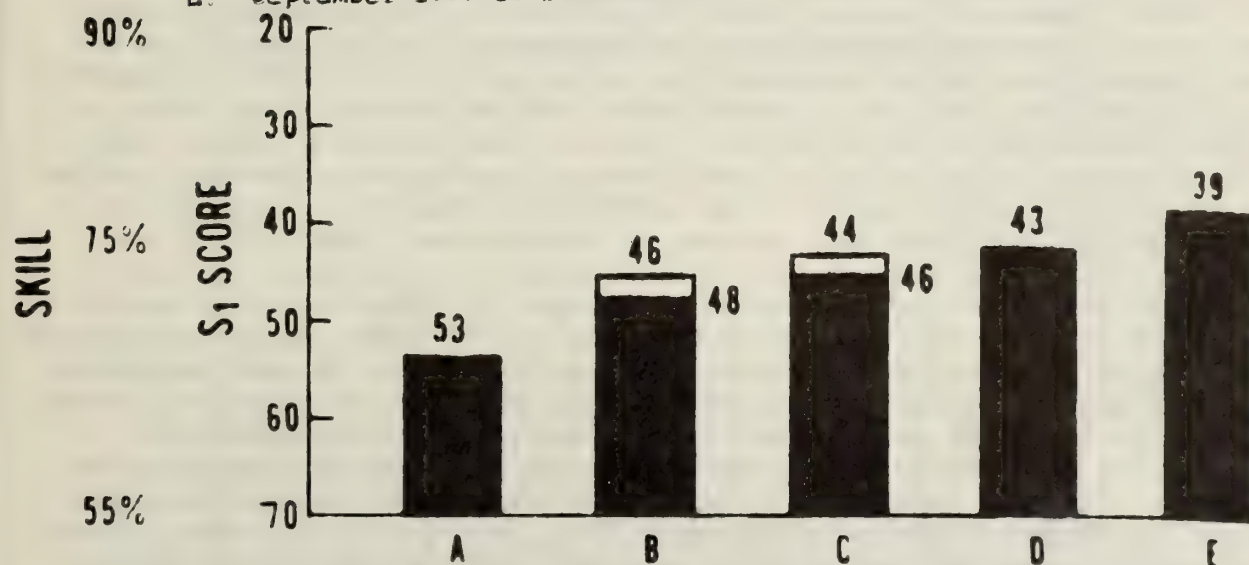


Fig. 2. Same as Fig. 1 but for NMC average 36 hour 500 mb.  $S_1$  score.

### 3. THE NATURE OF THE AVAILABLE RESULTS

In this paper, we do not attempt to make a general survey of the literature dealing with sun-weather research, and most of the individual



results in fact have been excluded. This is not meant to diminish the importance of those results not included here, but since we are dealing with the topic of the possible impact of sun-weather research on forecasting, we have been more concerned with the type of input that is available rather than covering all the possible types of input that are available.

We begin by considering the result of Wilcox et al. (1973). This, along with the result of Roberts and Olson (1973), is the most often quoted and best substantiated result in the literature. Wilcox et al. found that the vorticity area index (VAI) decreased by approximately 10% one day after a solar boundary crossing. The curve for the northern hemispheric VAI is shown in Figure 3. Since the VAI sums the area in the northern hemisphere where the vorticity exceeds a certain discriminator level, in this case  $20 \times 10^{-5} \text{ sec}^{-1}$ , this indicates that on the average, the area covered by low pressure will decrease by  $\sim 10\%$  following the passage of a solar magnetic sector boundary passage at the Earth. Since SSB crossings occur on the average once a week, this seems as if it could be an important result for forecast operations especially since the trigger is supplied by a solar input, something which is not considered within the normal realm of meteorology. However, we should emphasize the things that the forecaster would need to know in order to use this information. The fact that the low pressure areal coverage will decrease will be important since this could mean the difference between a "snow" and a "no snow" forecast if the forecast location is on the fringes of a cyclonic area in the winter. The important parameters that we don't know are whether or not the movement of a low, as well as its areal extent, is affected, whether lows in one area of the northern hemisphere are affected more than lows in other areas, and the nature of the effect in regions with higher vorticity levels. Knowing that the low pressure area will decrease in size following a SSB passage is valuable information, unless the low at the same time moves in an unpredictable manner. If the low is smaller but happens to be many degrees of latitude away from where it was expected to be, this information is not helpful to the forecaster. Also, we need to know if the effect varies in strength from one region to the next so that we can find the areas where knowledge of the effect would be important. Finally, the effect on the vorticity gradient, if any, is in some sense more important than the effect on the area in which the vorticity exceeds a certain value since it is the positive vorticity advection which determines the magnitude of the upward vertical velocities and therefore the precipitation patterns. Although these are all serious deficiencies in our understanding of the sun-weather effect at the present time, they are certainly not outside the scope of understanding which can be gained with the appropriate studies.

Roberts and Olson (1973) found that the vorticity area index associated with a specific trough increased following sharp rises in geomagnetic activity. The study by Wilcox et al. (1973) is considered to be more important in the field of sun-weather research since it directly linked the tropospheric circulation to solar activity and eliminated the possibility that the sun-weather effect was merely indicative of upper and lower atmosphere coupling without any dependence on the solar input, but the Roberts and Olson study is somewhat better suited for applications to forecasting. They followed individual troughs as they progressed from the Gulf of Alaska across the North American continent and found that on the average their size increased if they had been in the Gulf of Alaska shortly after high geomagnetic activity occurred. Wilcox and Duffy (1978) have



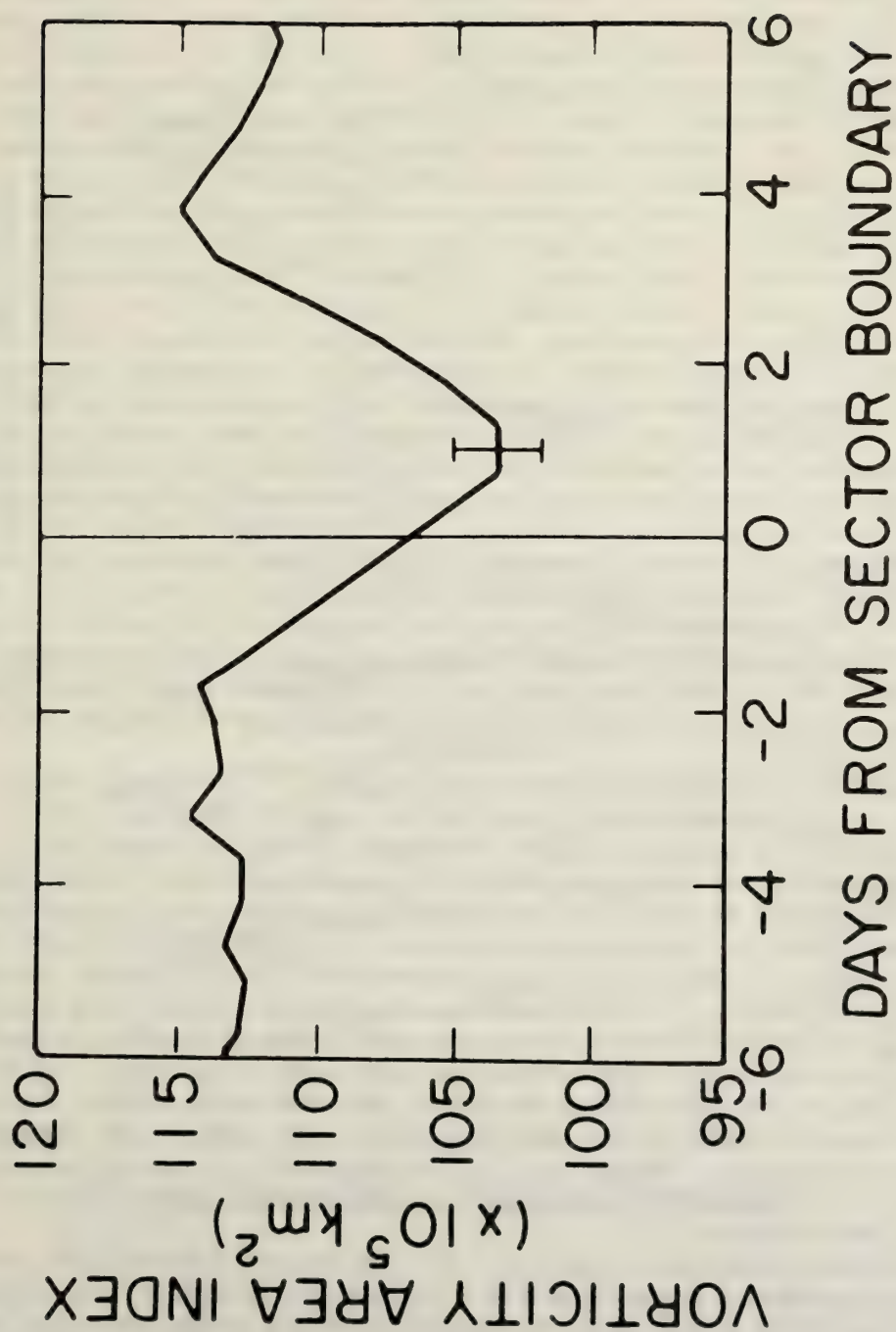


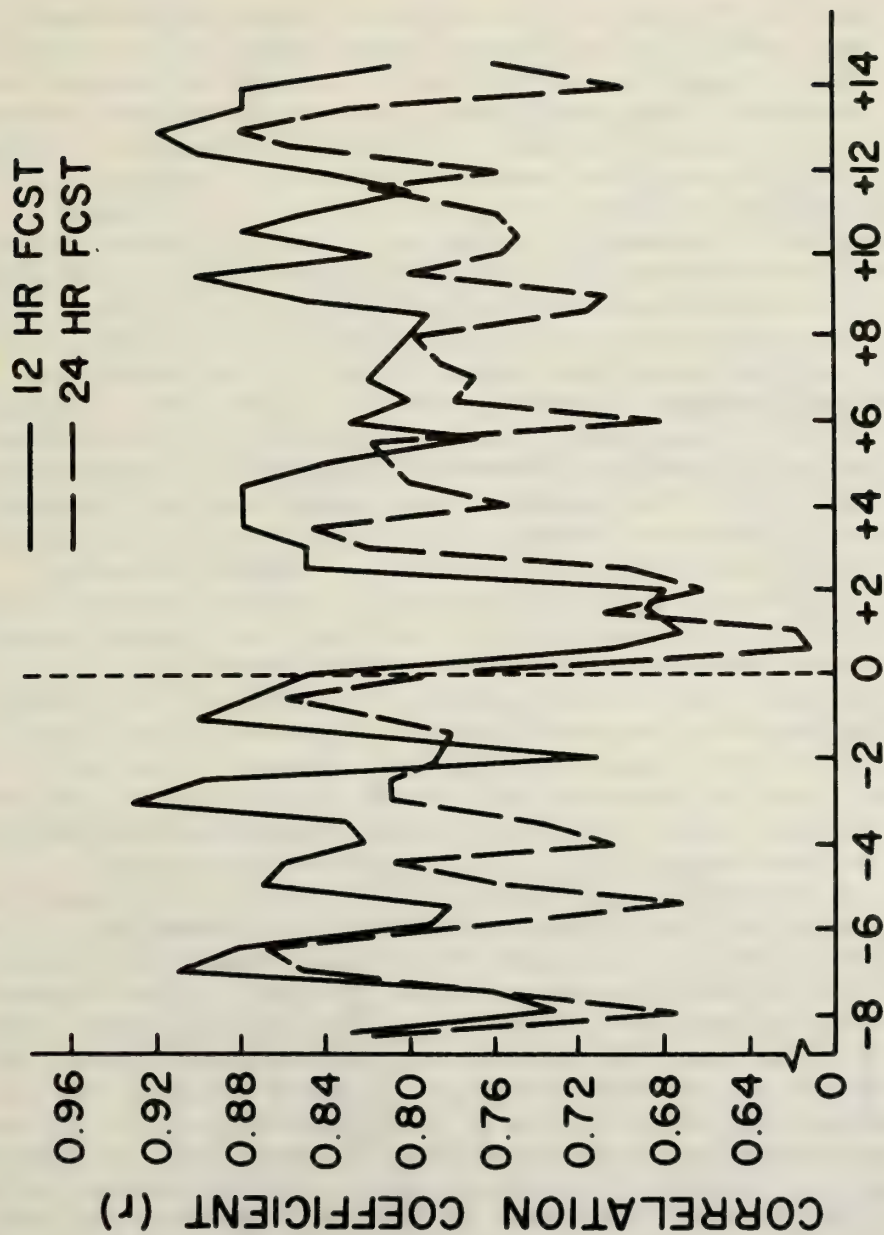
Fig. 3. Superposed epoch analysis of hemispheric vorticity area index.

reanalyzed the Roberts and Olson data and have found that there is an effect on trough size related to the direction of the interplanetary magnetic field at the time when the trough is in the Gulf of Alaska. The uncertainties in these studies are the same as those cited earlier, namely, possible effects relating to the movement of the troughs and the effect on the vorticity gradient.

Larsen and Kelley (1977) investigated the effect of SSB crossings on the ability of the Limited-area Fine Mesh numerical model to forecast the vorticity area index 12 and 24 hours later. The LFM is the best of the numerical models produced by NMC (Ramage, 1976). It has a grid size of 190 km which is half of the grid size of the other two models which are run routinely. It is a primitive equation model which takes into account moisture and radiation among other things. The results of the study are shown in Figure 4. The two curves represent the correlation coefficient between the observed and forecasted values of the vorticity area index for the 12 and 24 hour forecasts. The curves indicate that the correlations drop by 0.15 from their normal mean values in the period from 1 to 3 days after a solar magnetic sector boundary crossing. We emphasize that these curves represent correlation coefficients between the observed and forecasted VAI, and not a skill score. The problems associated with defining appropriate skill scores for the forecast quantities of interest would take more space to outline than is suitable here, but an excellent brief analysis can be found in Sommerville (1976). Figure 4 shows that whatever the mechanism for the sun-weather effect may be, it is something which is not accounted for even in the physics of the best numerical forecast product we have. The possibilities which come to mind are that the mechanism is associated with some of the smaller scale wave or cloud phenomena that are not handled by the model, or possibly a planetary scale influence from the layers of the atmosphere above those handled by the LFM. Since the study only covered the years from 1971 to 1974, it would be profitable to extend this type of study to a longer period and other models.

One difficulty associated with doing studies of forecast verifications as they relate to the sun-weather effect is the lack of available data. Since forecasters have not generally been concerned with this topic, skill scores have been calculated as averages over a one month period or even yearly averages. This sort of data is not useful in dealing with effects that occur several times in the averaging period. Sanders (1973) reported on the detailed records of forecasting skill made at MIT for forecasts verified at Logan Airport in Boston. These records extend back to 1967 and might provide an excellent opportunity for the type of study that is needed to evaluate the possible impact of the sun-weather effect on forecasting. The numerical forecast products generally only serve as guidance material for forecasts. The final forecast decision is usually made by the individual forecaster based on this guidance. Therefore, it is important to ask if the mis-guidance provided by the numerical product has any significant effect on the final man-machine mix forecast.

One problem inherent in applying the results of sun-weather research to forecasting procedures stems from the statistical nature of the investigations. Any statistical result will only be accepted if the error bars and hence the variance of the distribution is small. Hence, this automatically precludes any effects in which the magnitude of the perturbations induced in the tropospheric flow are highly dependent on other factors related to pre-existing conditions in the atmosphere. For instance, it might be found that



#### DAYS FROM SOLAR SECTOR BOUNDARY CROSSING

Fig. 4. Correlation coefficients between the 12 hour (solid line) and 24 hour (dashed line) forecasts and the observed vorticity area index.



if the sun-weather results were applied to operational forecasts, the magnitude of the effect would vary a great deal from one time to the next. This is a problem which will probably only be resolved once the mechanism is understood.

#### 4. POSSIBLE FORECAST IMPROVEMENTS

Having examined what we know and the type of things that we need to know in order to make our results applicable, this brings us to the type and magnitude of possible improvements in local forecasts which might be realized from our knowledge. In other words, if we apply all that we know in the form that we need to know it, will this make a worthwhile difference in the quality of the forecasts that are now available. To answer this question, we can start by asking what is the quality of the forecast product that we now have. Generally, the skill of various forecasts is measured by the  $S_1$  score which is the root mean square error of the forecast field. The trend for the forecast skill from 1947 to 1972 is shown in Figure 5. The upper curve shows the predictive skill for the sea level pressure gradient, and the lower curve shows the skill for predictions of the 500 mb. geopotential height (Miyakoda, 1974). Shuman (1972) states that an  $S_1$  score of 0.30 for the sea level prediction and 0.20 for the 500 mb. prediction are essentially perfect. Scores of 0.80 and 0.70 respectively are essentially worthless. The downward trend shows that the forecast skill has improved, but that forecasts are still far from perfect. The surface forecast suffers from greater inaccuracies than the 500 mb. forecast.

Sanders (1973) has studied the forecast skill of the staff and students at MIT over a six year period. He has concluded that the forecast skill for precipitation drops to 10% within  $2\frac{1}{2}$  days and the forecast skill for temperature drops to 10% after a 4 day period. He concludes that this is due to the large influence of mesoscale phenomena in determining these parameters which are of vital importance for many sectors of society. This serves to point out that the standard skill score evaluations of the predicted 500 mb. height field and surface pressure are not necessarily indicative of skill in forecasting associated with other atmospheric parameters.

The question of atmospheric predictability is one of basic importance in evaluating the impact of the sun-weather effect of forecasting. It is possible that the solar input may provide the means by which the accuracy of the forecast products can be extended to longer periods. If the forecast errors beyond one or two days are in part due to this missing input, then it should be possible to increase the validity of the forecasts beyond this period. The predictability of the atmosphere is limited in time due to our imperfect knowledge of the initial state of the atmosphere at the forecast time as well as uncertainties in the formulation of the equations describing the atmosphere (Lorenz, 1965). If the input due to the sun-weather effect is added to our modelling of the atmosphere, it will only increase the forecast skill out to the limit of the range of time in which the atmosphere is predictable. One way to clarify this concept is by considering that the initial data that the model is provided with will have errors in it, and these errors will have a certain variance. The effect of the errors is exactly the same as that of the real atmospheric waves that exist in the initial data field. They will propagate, grow, dissipate, and interact with other scales of motion. At some time or another, these "error" waves will dominate the model

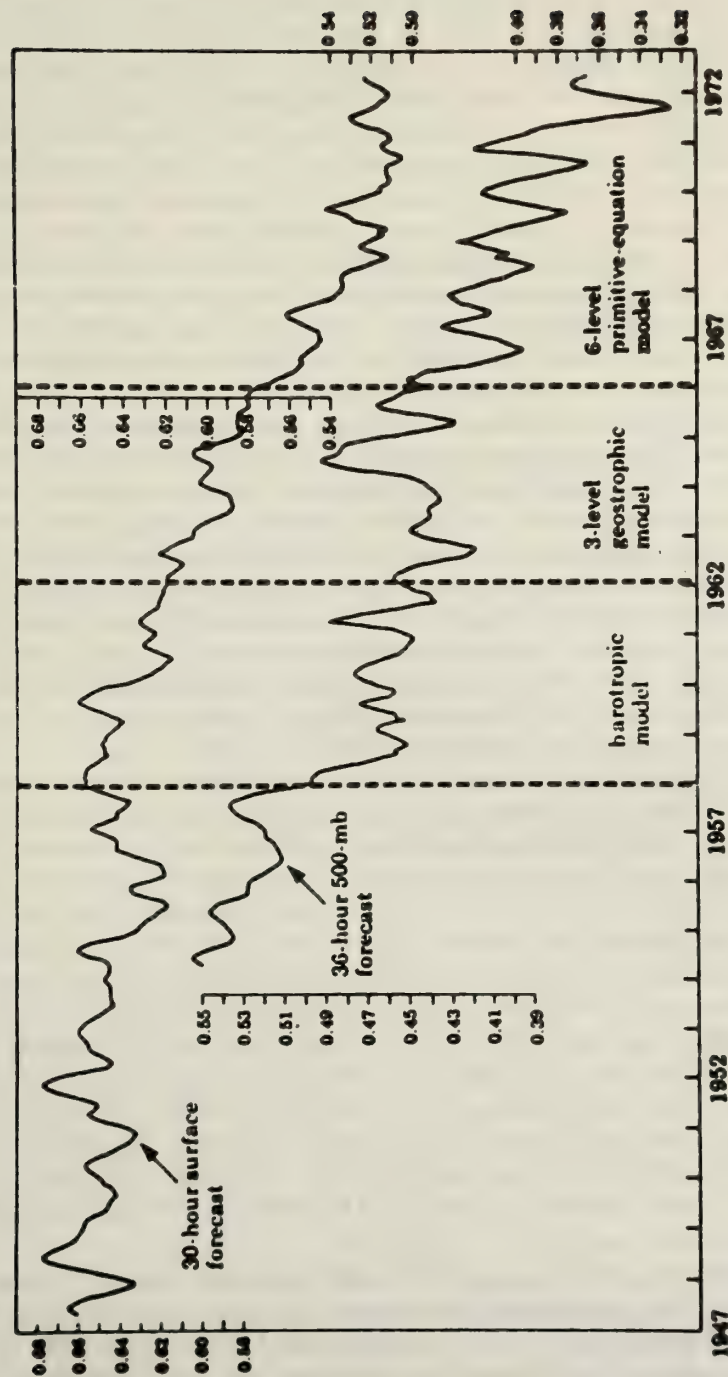


Fig. 5. Progress in numerical weather prediction from 1947 to 1972 as shown by the  $S_1$  score.



output and this limits the time range of predictability of the atmosphere. If this predictability time scale is of the same order as the time from observing an event on the sun to its arrival at the Earth, then we cannot hope to extend the forecast period by invoking knowledge of solar activity as an indicator of atmospheric activity beyond the limit of predictability. The solar wind stream takes four to five days to reach the Earth on the average although it can be less during periods of high solar activity. It might be postulated that if the solar effect on the atmosphere is significant, this four to five day lag will give us an advantage in extending the forecast range. Also, the solar sector structure is known to have some periodicity since the sun's rotation sweeps these sectors past the Earth (Svalgaard, 1972).

Predictability studies have involved one of three techniques. The first integrates a set of equations describing the atmosphere with two different sets of initial conditions. The divergence of the two solutions is determined and the doubling time of the errors calculated. The second approach examines the rate at which similar weather patterns diverge statistically. The last approach couches the equation for the atmosphere in spectral form, and the spectral distribution of the errors is calculated. Smagorinski (1969) has examined the problem using the NOAA Geophysical Fluid Dynamics Laboratory primitive equation general circulation model. An error with a standard deviation of  $0.5^{\circ}\text{C}$  was input at all levels. The doubling time of the errors was found to be 2.5 days. The statistical approach has been used by Lorenz (1969). He used five years of data and calculated the RMS error between similar weather patterns occurring in the same month of the year. The doubling time of these errors was also found to be 2.5 days. This doubling time leads to the conclusion that the larger scales are only predictable over a range of several weeks at most.

Lorenz (1969b) and Robinson (1971) have investigated the predictability problem using the spectral models. In the spectral models, it is assumed that all the error exists at the smaller scales which cannot be resolved in the initial data field. These errors at the smaller scales interact with the larger scales and eventually destroy the characteristics of the larger scales through turbulent diffusion. The two studies concluded that cumulus scales were predictable for a period of about one hour, synoptic scale motions for a few days, and the larger scales for about 17 days.

These results indicate that if the sun-weather research is to have any impact on forecasting on the longer time scales, it will have to be at the larger scales in the atmosphere. If the solar effect occurs at the synoptic scale, we cannot hope to extend the forecasts much beyond the one to two days which are presently handled operationally, although the skill of the forecasts might improve. There seems to be little hope of any improvement at the shortest scales.

Robert (1976) has evaluated the various sources of error for a 500 mb 24 hour forecast. He estimates that the sources of error in the forecast are as follows in Table 1. The percentages listed in Table 1 are estimates of the percentage error variance contributed by each entry. The percentages in the table can be interpreted as follows. For item 1 for instance, the mean square difference between the geopotential height field forecast with a 400 km grid and that forecast with a 200km grid is 35% of the mean square difference between the forecasted and verification geopotential height fields. The fact that the percentages add up to more than 100% is attributable to the fact that the errors are correlated. When the



TABLE 1

1. Horizontal resolution: 400 km vs. 200 km	35%
2. Mountains vs. no mountains	27%
3. Precipitation vs. no precipitation	18%
4. Five levels vs. high vertical resolution	15%
5. Inadequate observing system	13%
6. Horizontal eddy viscosity vs. no viscosity	11%
7. Surface drag vs. no surface drag	8%
8. Refinements of radiative, viscosity terms, inclusion of detailed mountain profiles, etc.	4% each

individual cases are examined, certain errors from other categories will appear. Miyakoda (1974) has made the same sort of analysis, and he concluded that the inadequate observing system should be second on the list in importance. If the evaluations of the errors by Robert are correct, this indicates that the inclusion of the sun-weather effect in the models at the present time should be carefully weighed in terms of the possible improvements that could be expected. A detailed investigation of the error created by its omission is needed so that it can be ranged along-side the other sources of error. The result of Wilcox et al. (1973) indicates that on the average, the vorticity isopleth will enclose an area that is 10% smaller after the passage of a solar magnetic sector boundary crossing. If this is representative of the error generated by the exclusion of the sun-weather effect from the model dynamics, then it would have to be placed somewhere near fifth or sixth place on the list in Table 1. On the other hand, Hines and Halevy (1977) concluded after analyzing the same data used by Wilcox et al. that the effect does not change the atmospheric events which would occur anyway, but rather changes the timing of the occurrence of those events. If this is the case, then the error could be much larger than 10% since the forecast skill depends both on predicting the magnitude of the atmospheric gradients as well as their position. This is an area that needs further study.

## 5. POSSIBLE IMPACT ON PREDICTIONS ON A CLIMATIC TIME SCALE.

In the previous section we found that the theoretical limits of predictability increase as the scale size of the phenomena that we are interested in increases. Therefore, the longest-range forecasts should only be possible for the very largest scales. Yet, these long-range forecasts are not possible today (Haltiner, 1975). Schneider (1976) defines climatic changes as those which occur over a period of 90 days or longer. He finds that although the predictability of the atmosphere is of the order of several weeks, that it should still be possible to predict longer term means. It would not be possible to predict the mean temperature at a certain location on a specific day 90 days in advance, but it should be possible to predict the mean temperature for the next 90 day period. The problem that arises in making these long term predictions of mean quantities is determining the periodicities that are important. Kutzbach and Bryson (1974) found that the climatic record over a period of 10,000 years pieced together from various sources shows randomness for periods of ten years or less. If these variations are truly random, then there is little hope of formulating predictions for a time scale of less than ten years. If significant fluctuations exist

in the climatic means due to variations of solar activity, then these variations would serve as a basis for long-term forecasting. King (1975) has summarized many of the results of sun-weather research relating to the 11 and 22 year solar cycles. Pittcock (1978) has criticized many of the results derived from these studies, but he has also advanced some criteria for future studies to eliminate some of the vagaries in the earlier ones. Pittcock has recommended that new results be evaluated by means of their predictive skill. If climatic fluctuations with periodicities related to the periodicity of the solar cycle are found, then the ability of this result to predict variations within the next 11 or 22 year cycle would serve as a check on the validity of the result.

It may be that solar cycle variations have a significant effect on the atmosphere in the long-term mean. If this is the case, then these variations could serve as the basis for long-term forecasts. The obstacle that we encounter in trying to apply the results of sun-weather research dealing with solar cycle variations is the long verification time. Predictive schemes formulated today could be evaluated over the next solar cycle and applied in the one that follows. This is not encouraging in terms of improving long-term forecasting within the next few years, although it may be the area in which sun-weather research will have the greatest impact.

## 6. CONCLUSION

Our evaluation of the impact of sun weather research on forecasting suggests that at the very shortest spatial scales (mesoscale and smaller phenomena) the impact will be negligible both in terms of creating more accurate forecasts or extending the period of time over which the forecasts will remain valid. At the synoptic scale the limit of atmospheric predictability is theoretically greater than the period for which we now have accurate forecasts. It is possible that inclusion of the results of sun-weather research in the routine forecast procedures for the synoptic scale will improve the quality of the forecasts somewhat, but the gain in forecast skill has to be weighed against the magnitude of other errors that presently exist in the models. At the hemispheric scale, the possibility that numerical forecasts could be extended to the limit of several weeks is not excluded. In this case our aim would be to predict average circulation features over this time range rather than the instantaneous weather pattern on any given day.

The greatest impact of sun-weather research might well be in climatic forecasting if the atmospheric variations related to solar cycle variations are found to be significant and stand up under forecast verification tests. At the present time we have no climatic prediction scheme. The results of sun-weather research could serve as the basis for such a scheme.



## REFERENCES

- Bengtsson, L. (1976): Problems and prospects for medium range weather forecasts. In: Weather Forecasting and Weather Forecasts: Models, Systems, and Users, Vol. 2. Published by NCAR, NCAR/CQ -5 + 1976-ASP.
- Cooley, D.S., and R.G. Derouin (1972): Long term verification trends of forecasts by the National Weather Service, NOAA Technical Memorandum NWS FCST - 18, U.S. Govt. Printing Office, Washington, D.C.
- Haltiner, G.J., and R.T. Williams (1975): Some recent advances in numerical weather predictions. Mon. Wea. Rev., 103.
- Hines, C.O., and I. Halevy (1977): On the reality and nature of a certain sun-weather correlation. J. Atmos. Sci., 43.
- King, J.W. (1975): Sun-weather relationships. Aeronautics and Astronautics, 13.
- Kutzbach, J.E., and R.A. Bryson (1974): Variance spectrum of Holocene climatic fluctuations in the North Atlantic sector. J. Atmos. Sci., 31.
- Larsen, M.F., and M.C. Kelley (1977): A study of an observed and forecasted meteorological index and its relation to the interplanetary magnetic field. Geophys. Res. Let. 4.
- Lorenz, E.N. (1965): A study of the predictability of a 28-variable atmospheric model. Tellus, 17.
- Lorenz, E.N. (1969): The predictability of a flow which possesses many scales of motion. Tellus, 21.
- Lorenz, E.N. (1969b): Atmospheric predictability as revealed by naturally occurring analogues. J. Atmos. Sci., 26.
- Miyakoda, K. (1974): Numerical weather prediction. Amer. Sci., 62.
- Pittock, A.B. (1978): A critical look at long-term sun-weather relationships. Rev. Geophys. and Spa. Phys., 16.
- Ramage, C.S. (1976): Prognoses for weather forecasting. Bull. Amer. Meteor. Soc., 57.
- Robert, A. (1976): Sensitivity experiments for the development of NWP models. Proceedings of the (Eleventh) Stanstead Seminar. Publication in Meteorology No. 114, McGill University, Montreal, Canada.
- Roberts, W.O. and R.H. Olson (1973): Geomagnetic storms and wintertime 300 mb. trough development in the North Pacific-North America area. J. Atmos. Sci., 30.



- Robinson, G.D. (1971): The predictability of a dissipative flow. Quart. J. Roy. Meteor. Soc., 97.
- Sanders, F. (1973): Skill in forecasting daily temperature and precipitation, some experimental results. Bull. Amer. Meteor. Soc., 54.
- Schneider, S.H. (1976): The response to climate change: the significance of different time scales. In: Living with Climatic Change, Phase II. The Mitre Corp., McLean, Virginia.
- Shuman, F.G. (1972): The research and development program at the National Meteorological Center. Office Notes 72, NMC, NOAA.
- Smagorinsky, J. (1969): Problems and promises of deterministic extended range forecasting. Bull. Amer. Meteor. Soc., 50.
- Sommerville, R.C.J. (1976): Model verification. In: Weather Forecasting and Weather Forecasts: Models, Systems, and Users, Vol. 2. Published by NCAR, NCAR/CQ -5 + 1976 -ASP.
- Svalgaard, L. (1972): Interplanetary magnetic sector structure, 1926-1971, J. Geophys. Res., 77.
- Wilcox, J.M., P.H. Scherrer, L. Svalgaard, W.O. Roberts, and R.H. Olson (1973): Solar magnetic sector structure: relation to circulation of the earth's atmosphere. Science, 180.
- Wilcox, J.M. and P. Duthie (1978): Data presented at the Symposium/Workshop on Solar-Terrestrial Influences on Climate and Weather, Columbus, Ohio.

# PLANETARY WAVES AND SUN-WEATHER EFFECTS\*

H. Volland  
Space Physics Research Laboratory  
Department of Atmospheric and Oceanic Science  
The University of Michigan  
Ann Arbor, Michigan 48109, U.S.A.\*\*

J. Schäfer  
Radioastronomical Institute  
University of Bonn  
5300 Bonn, W. Germany

In order to interpret sun-weather effects, one has to understand planetary wave propagation. In this paper, a brief outline about the theory of planetary waves is given and a review about recent work on space-time analysis, mainly at the 500 mbar pressure level, is presented. This analysis gives evidence that broad spectral bands of two types of waves exist within the troposphere: ultralong waves with zonal wave numbers  $m \leq 4$  and periods  $\tau \gtrsim 5$  days, propagating mainly to the west, and synoptic scale waves with  $m \gtrsim 3$  and  $\tau \lesssim 10$  days, propagating mainly to the east. These waves are generated by internal turbulent processes within the atmosphere and are quasi-persistent with lifetimes of several periods. It is shown that solar activity cannot generate planetary waves of significant amplitudes, and that the observed "sun-weather effects" can be interpreted within the framework of these internally generated planetary waves without any trigger mechanism from outside the atmosphere. It is suggested that a better knowledge of these persistent ultralong waves may help to improve long range weather forecasts.

---

\* Invited paper; Proceedings and Workshop on Solar-Terrestrial Predictions, April 23-27, 1979, Boulder, Colorado.

\*\* Permanent Address: Radioastronomical Institute, University of Bonn, 5300 Bonn, W. Germany.

## 1. INTRODUCTION

What we call weather, is the irregular transport of heat and angular momentum through the atmosphere by winds and eddies, mainly from lower to higher latitudes. Planetary waves play, hereby, a central role in global scale weather and climate. These waves possess wavelengths of 1000 km or longer, so that the spherical waveguide characteristics of the rotating earth influence significantly their propagation conditions. If  $m$  is the zonal wave number and  $\tau$  (in days) is the period of these waves,  $a$  is the earth's radius, and  $f \approx 2\pi/(1 \text{ day})$  is the Coriolis parameter at middle latitudes, then the characteristic length of the waves is  $L \approx a/m$ , the characteristic phase velocity is  $c \approx 2\pi a/(m\tau)$ , and the Rossby number becomes  $Ro \approx c/(fL) \approx 1/\tau$ .

Observations and theory point to a basic distinction between two types of waves: synoptic scale waves with zonal wave numbers  $3 \lesssim m \lesssim 10$  and typical periods of  $\tau \lesssim 10$  days, propagating mainly eastward, and ultralong waves with  $m \lesssim 4$  and typical periods  $\tau \gtrsim 5$  days, propagating mainly westward. Tidal waves with  $m \approx 3$  and  $\tau \approx 1$  day belong to the second type. For the first type of motion, the magnitude of horizontal divergence is small compared to the vertical component of vorticity. For the second type, the horizontal divergence cannot be neglected [Phillips, 1963].

The synoptic scale waves arise from instabilities in the mean westerly flow. Their properties are fairly well understood [e.g., Holton, 1975]. Efforts to formulate prediction models to forecast ultralong waves were not so successful [Kasahara, 1976; Hayashi and Golder, 1977]. A proper understanding of the ultralong waves is, however, an essential component in any theoretical picture of the general circulation. Observational studies have shown that more than half the total poleward heat transport in northern middle latitudes in winter is accomplished by wave numbers 1 to 4 with wave number two tending to predominate [Wiin-Nielsen, 1964; Wiin-Nielsen et al., 1963].

Knowledge of the generation and the propagation of planetary waves is also the key in understanding the various solar activity-weather (or simply "sun-weather") effects. According to Wilcox [1975], many short term sun-weather effects reported in the literature have a few common features in spite of their diversities and their controversial aspects: they tend to be most pronounced during the winter season, and the meteorological responses over the oceans tend to be opposite to the responses over the continents. These features, indeed, reflect properties of planetary waves because it is well known that extratropical planetary wave propagation is optimum during the winter [Charney and Drazin, 1961], and that a strong wave number two component exists which reflects the land-sea coverage in the northern hemisphere [van Loon et al., 1973].

The only possible response of the atmosphere on a synoptic or a global scale to any kind of disturbance is via planetary waves. Therefore, any hypothesis which tries to explain sun-weather effects must ultimately specify the generation of planetary waves. A popular hypothesis is that solar activity may change the electric conductivity of the air which influences the global geoelectric field and thus the thunderstorm activity [Markson, 1978]. How then planetary waves are generated and on what time scale,



remains an open question.

To date, there is still no general agreement about the reality of sun-weather effects. The various long term effects involving the single and the double sunspot cycle have been critically reviewed by Pittcock [1978]. The short term effects with time scales from days to several weeks in connection with sector boundary crossings of the interplanetary magnetic field, sunspots, or solar flares, have received much attention recently. In particular, the Roberts-Wilcox relation between sector boundary crossings and the vorticity area index [e.g., Wilcox et al., 1974] has been positively tested by Hines and Halevy [1977] and by Larsen and Kelley [1977]. Larsen and Kelley [1977] found a deterioration of the 12 h and the 24 hour weather forecasts about two days after the sector boundary crossings. On the other hand, Bhatnagar and Jakobsson [1978] were not able to detect a correlation between sector boundary crossings and pressure data at the 500 mbar level. King et al. [1977] found a 27.5 day periodicity in the height of the 500 mbar pressure level which they related to rigidly rotating active longitudes on the sun. They claim that these periodicities are predominantly standing waves with amplitudes of several geopotential decameters, sufficiently large even to block weather systems.

Before one starts to include findings such as those of Larsen and Kelley [1977] and King et al. [1977] into meteorological weather forecasts, one has to understand at least the basic mechanism involved. In this paper we shall try this in two steps: in the next section, we shall estimate from first principles the amplitudes of such waves which can possibly be expected from a variable sun. In Section 3 we shall review some basic knowledge about planetary wave propagation. In Section 4, we shall outline recent work on space-time analysis of the 500 mbar level [Willson, 1975; Fraederich and Böttger, 1978; Schäfer, 1979] which involves the whole spectrum of planetary waves from about 2 to 40 days.

Unfortunately the conclusions drawn both from theory and data analysis will turn out to be rather negative with respect to "sun-weather effects". The atmosphere is able to generate internally a broad spectrum of quasi-persistent planetary waves, and the observed effects can be interpreted within that framework. There is, therefore, no need for any trigger mechanism from outside the atmosphere.

## 2. ESTIMATE OF PRESSURE DISTURBANCES GENERATED BY SOLAR ACTIVITY

Present estimates of the short term variability of the total solar radiation (the solar constant) put an upper limit of 0.1% in the variation of the visible light [Livingston, 1978]. The highly variable solar XUV-component which so dramatically disturbs the upper atmosphere is only  $10^{-6}$  of the solar constant [Willis, 1976]. The electric energy involved in the thunderstorms and in the geoelectric field is about  $10^{-3}$  of the kinetic energy of the atmosphere. Even if one imagines some trigger mechanism, e.g., albedo changes due to a changing ozone content during solar activity, the solar radiation reaching the ground is still constant within about 0.1% [Foukal et al., 1977]. Therefore, that part of the energy input from the sun either directly or indirectly available in connection with solar activity is at most  $S = 1 \text{ W/m}^2$ .

In order to estimate the reaction of the atmosphere to that tiny heat source, we use the first law of thermodynamics which can be written as

$$\frac{\gamma}{\gamma-1} \frac{1}{T_o} \frac{dT}{dt} - \frac{1}{p_o} \frac{dp}{dt} = \frac{1}{R} \frac{Q}{T_o} \quad (1)$$

where  $\gamma$  is the ratio of the specific heats,  $T_o$ ,  $p_o$  are basic temperature and pressure at the ground,  $T$ ,  $p$  are temperature and pressure variation,  $R$  is the gas constant, and  $Q$  an external heat input. Assuming a constant heat input per unit mass within the troposphere, averaged over the sphere, it is

$$Q = \frac{S}{4 \rho_o H} \quad (2)$$

where  $\rho_o$  is the density at the ground,  $H$  the pressure scale height, and the factor four a geometry factor. Neglecting the temperature variation and assuming a periodic heat source of period  $\tau$ , the pressure amplitude at the ground due to that heat input is [Volland, 1977]

$$|p| = \frac{\tau S}{8\pi H} \quad (3)$$

Equation (3) shows that the reaction of the atmosphere to a thermal heat source does not only depend on the magnitude of  $S$ , but also on the period with which the source varies. The atmosphere, therefore, behaves like a low pass filter suppressing higher frequencies. Of course, the validity of equation (3) breaks down for larger periods due to dissipation effects. If one approximates thermal dissipation by Newtonian cooling,  $\tau$  in (3) has to be replaced by

$$\tau' = \frac{\tau \tau_o}{\tau_o + \tau} \quad (4)$$

with  $\tau_o \approx 1.3 \times 10^7$  s = 150 days [Dickinson, 1973].

We can test the usefulness of equation (3) inserting approximate numbers for the tidal waves. For the diurnal tide, it is  $\tau = 1$  day,  $H \approx 8$  km, and  $S \approx \alpha S_o \approx 250$  W/m<sup>2</sup> ( $\alpha$  is that fraction of the solar constant  $S_o$  which is directly absorbed within the atmosphere). Thus  $|p| \approx 1$  mbar. The observed diurnal tide at the equator is 0.6 mbar [Chapman and Lindzen, 1970], therefore within a factor of two the number given by equation (3). As it is well known [e.g., Chapman and Lindzen, 1970], it is already difficult at the ground to filter the thermally driven diurnal tidal waves out of the interdiurnal meteorological noise which has amplitudes of 10 - 20 mbar at middle latitudes.

If one expects that the sun acts like a rotating light house with a period of 27 days and an amplitude of  $S = 1$  W/m<sup>2</sup>, one arrives at  $|p| = 0.1$



mbar. Thus, the reduction in S by a factor of 250 relative to the diurnal tide is partly compensated by the increase of the period by a factor of 27. That pressure amplitude is of the same order of magnitude as the amplitude of the semidiurnal lunar tide which is 60  $\mu$ bar near the equator. Chapman [e.g., Chapman and Lindzen, 1970] needed pressure data of 60 years in order to find that small signal, in spite of the fact that the lunar tidal force is well known and behaves quite regularly. It seems very improbable, therefore, to detect out of a much smaller data base a comparably small signal generated by a source which is noisy itself.

To interpret the 27 day period in the pressure as a tidal effect due to the orbital motion of the moon is also impossible because the lunar monthly tidal force is more than one order of magnitude smaller than the semidiurnal lunar tidal force, and the low pass filter effect does not work in the case of a momentum force.

### 3. BASIC FEATURES OF PLANETARY WAVE PROPAGATION

#### 3.1 Laplace's Tidal Equations

In the last section, we have shown that the active sun cannot generate planetary waves with significant amplitudes. In this section we want to outline briefly the theory of planetary wave propagation which will help to understand the various observations of "sun-weather effects".

The most convenient way to describe planetary wave propagation is in terms of linear tidal theory. The general idea is that the waves are small perturbations from a basic state and that no feedback occurs between waves and basic state. If the parameters of the basic state depend only on height - with the exception of the prevailing westerly wind - it is possible to obtain solutions of the hydrodynamic equations which are sums of products of functions of height  $z$ , of latitude  $\phi$  and of longitude  $\lambda$ ; e.g., for the pressure it is [Chapman and Lindzen, 1970; Volland and Mayr, 1977]

$$p = \sum_{n,m} p_n^m(z) \Theta_n^m(\phi) e^{im\lambda - i\omega t} \quad (5)$$

Each term in the series represents an individual and independent wave. The waves propagate to the east (west) if the frequency  $\omega$  is positive (negative).

The horizontal velocity can be separated into a curl free component  $\Phi_n^m$  and a source free component  $\Psi_n^m$ :

$$\underline{U} = a \sum_{n,m} \{ -U_n^m(z) \nabla \Phi_n^m(\phi) + \bar{U}_n^m(z) \nabla \times \Psi_n^m(\phi) \hat{r} \} e^{im\lambda - i\omega t} \quad (6)$$



with  $\hat{r}$  the unit vector in r-direction. Moreover, the separation of the variables implies a proportionality between the horizontal wind  $\underline{U}$  and the pressure  $p$ :

$$\nabla \cdot \underline{U} + \frac{1}{\rho_0 g h} \frac{Dp}{Dt} = 0 \quad (7)$$

where  $h$  is a separation constant to be determined later,  $g$  the gravitational acceleration and the total time derivative is

$$\frac{D}{Dt} \approx -i\omega + \frac{imU_0}{a} = -i\omega_D$$

if the atmosphere superrotates rigidly with velocity  $U_0 \cos\phi$ .

The equation of horizontal momentum is

$$\frac{D\underline{U}}{Dt} + 2\Omega \hat{r} \times \underline{U} \sin\phi + \frac{1}{\rho_0} \nabla_{\text{hor}} p = 0 \quad (8)$$

$\Omega = 2\pi/1$  day is the angular frequency of the earth's rotation.

Introducing (7) and (6) into (8) and applying the div and the curl operators, respectively, to (8) leads to Laplace's tidal equations [Longuet-Higgins, 1968]

$$\begin{aligned} i(\omega_D \nabla^2 - \frac{2m\Omega}{a}) \phi_n^m + 2\Omega \tilde{\nabla}^2 \psi_n^m - \frac{gh_n^m}{i\omega_D} \nabla^2 \nabla^2 \phi_n^m &= 0 \\ i(\omega_D \nabla^2 - \frac{2m\Omega}{a}) \psi_n^m - 2\Omega \tilde{\nabla}^2 \phi_n^m &= 0 \end{aligned} \quad (9)$$

with  $\nabla^2$  the horizontal Laplace operator, and

$$\tilde{\nabla}^2 = \sin\phi \nabla^2 + \frac{\cos\phi}{a^2} \frac{\partial}{\partial\phi}$$

It is

$$\omega_D = \omega - \frac{mU_o}{a} \quad (10)$$

a Doppler shifted frequency.

In meteorological terminology,  $-\nabla^2 \phi_n^m$  is the horizontal divergence, and  $-\nabla^2 \psi_n^m$  is the vertical component of the vorticity.

### 3.2 Rossby-Haurwitz Waves

The solutions of (9) become particularly simple in the case of zero horizontal divergence,  $\phi_n^m = 0$ . Then from (7) and (9) follows

$$\begin{aligned} \psi_n^m &= P_n^m \\ h_n^m &= \infty \end{aligned} \quad (11)$$

with  $P_n^m$  the spherical harmonics, and

$$U_o - c = \frac{2\Omega a}{n'(n' + 1)} \quad (12)$$

where  $c = a\omega/m$  is the phase velocity of the wave. Equation (12) is the famous Rossby formula [e.g., Holton, 1975] transformed into a spherical coordinate system. The latitude structure of the associated pressure, the function  $\Theta^m(\phi)$  in (5), is plotted in Figure 1 versus latitude for wave number  $m = n^2$ . These Rossby-Haurwitz waves have their largest amplitudes at middle and higher latitudes. The observed synoptic scale waves in the midlatitude westerlies with wave numbers  $m \gtrsim 4$  and periods of several days are of that type.

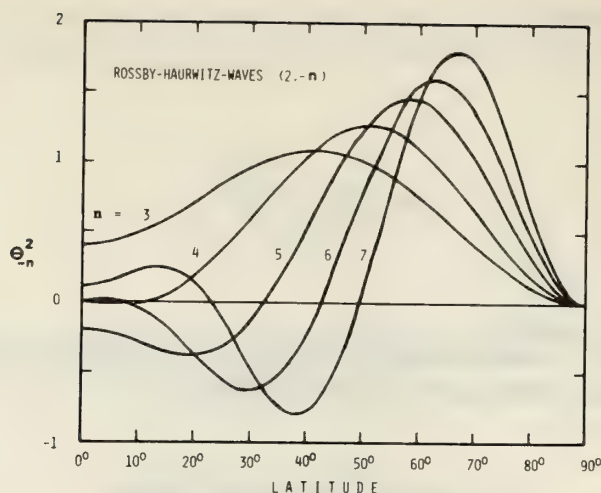


FIGURE 1: NORMALIZED HOUGH FUNCTIONS OF ROSSBY-HAURWITZ WAVES OF ZONAL WAVE NUMBER  $m = 2$  VERSUS LATITUDE. WAVES WITH ODD NUMBERS  $m-n$  ARE SYMMETRIC WITH RESPECT TO THE EQUATOR, WAVES WITH EVEN NUMBERS ARE ANTISYMMETRIC.

### 3.3 The Equivalent Depth

The rigorous solution of (9) constitutes an eigenvalue problem from which the eigenvalue  $h_n^m$ , sometimes called the equivalent depth, and the corresponding meridional structures of  $\Theta_n^m$ ,  $\Phi_n^m$ , and  $\Psi_n^m$  can be determined. In tidal theory,  $\Theta_n^m$  is called the Hough function.

Figure 2 shows the reciprocal equivalent depths  $1/h_n^m$  versus normalized frequency  $\omega/\Omega$  for wave number  $m = 2$  [Longuet-Higgins, 1968]. One notices two classes of waves. Waves of class I or gravitational waves (labelled by positive numbers  $n$ ) possess positive equivalent depths and exist in the whole frequency range. Waves of class II or rotational waves (labelled by negative numbers  $n$ ) possess positive and negative equivalent depths and disappear at frequencies  $|\omega| > 2\Omega$  or periods smaller than 12 hours. Evidently, the Rossby-Haurwitz waves belong to the westward propagating class II waves with  $1/h = 0$ . That occurs at the Rossby frequency

$$\omega_R = - \frac{2m\Omega}{|n|(|n| + 1)} \quad (13)$$

where  $n < 0$ . (Note that in (12), it is  $n' = -n$ ).

The height structure functions  $p_n^m$ ,  $U_n^m$  and  $\bar{U}_n^m$  in (5) and (6) must be determined from the remaining set of hydrodynamic and thermodynamic equations [e.g., Volland and Mayr, 1977]. They are connected with the meridional structure via the vertical wave number



$$k_n^m = \frac{1}{2H} \sqrt{\frac{4(\gamma-1)}{\gamma} \frac{H}{h_n^m} - 1} \quad (\text{if } T_o = \text{const}) \quad (14)$$

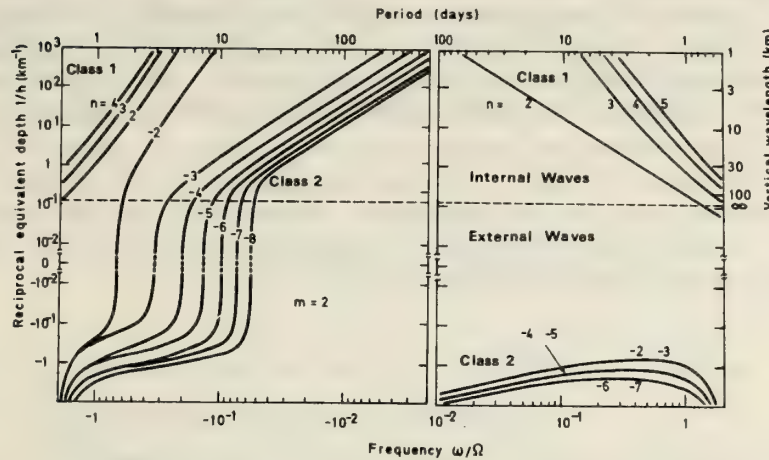


FIGURE 2: RECIPROCAL EQUIVALENT DEPTHS VERSUS NORMALIZED FREQUENCY FOR PLANETARY WAVES OF ZONAL WAVE NUMBER  $m = 2$ . SCALE ON THE RIGHT HAND SIDE GIVES VERTICAL WAVELENGTHS. WAVES WITH POSITIVE  $n$ : CLASS I WAVES OR GRAVITATIONAL WAVES; WITH NEGATIVE  $n$ : CLASS II WAVES OR ROTATIONAL WAVES. THE HORIZONTAL DASHED LINE SEPARATES THE INTERNAL WAVES WITH FINITE VERTICAL WAVELENGTH FROM THE EXTERNAL WAVES WITH INFINITELY LONG WAVELENGTH. NOTE THAT THE LOGARITHMIC ORDINATE INCLUDES POSITIVE AND NEGATIVE  $1/h$ .

Waves with  $k_n^m$ , real, are internal waves with finite vertical wavelengths

$$\lambda_n^m = 2\pi/k_n^m \quad (15)$$

The relative pressure amplitudes  $p_n^m/p_o$  of ascending free waves increase with altitude according to  $\exp\{z/(2H)\}$ . Waves with  $k_n^m$ , imaginary,  $(4(\gamma-1)H/\gamma/h_n^m < 1)$  are called external waves because they cannot transport wave energy upward. The relative pressure amplitudes of ascending free waves of that type decrease with altitude if  $|k_n^m| > 1/(2H)$ .

### 3.4 The Internal Class II Waves

The meridional structures of the internal waves of class I with periods greater than several days are confined to the equatorial regions; i.e., the Kelvin waves ( $n = m$ , east, in Figure 2) or the mixed Rossby-gravity waves

( $n = -m$ , west, in Figure 2) [Wallace, 1973]. Moreover, they possess vertical wavelengths smaller than 1 km (see right ordinate in Figure 2) and, therefore, are sensitive to destructive interference. The external class II waves with negative equivalent depths are suppressed. On the other hand, the meridional structure of the internal class II waves is similar to the structure of the Rossby-Haurwitz waves with maximum amplitudes at middle and higher latitudes (see Figure 1). Hence, extratropical wave propagation is dominated by the internal class II waves with negative frequencies. The negative frequencies which determine the wave modes and their meridional structures are the Doppler shifted frequencies in equation (10). This is Charney and Drazin's [1961] theory which states that vertical propagation of extratropical ultralong waves can occur only in the presence of westerly winds weaker than a critical value. According to Figure 2 and equations (12) and (13), that critical prevailing westerly wind velocity is

$$U_0 \approx a\Omega \left\{ \frac{1}{m\tau} + \frac{2}{|n|(|n| + 1)} \right\} \quad (16)$$

with  $\tau$  (in days) the period of the waves, which is positive for eastward traveling waves and negative for westward traveling waves.

The height structure of the internal class II waves depends on type and vertical profile of the forcing function, which may be either thermal or orographic forcing. It also depends strongly on the parameters of the basic state of the atmosphere. In particular, the height of the tropopause and/or the stratospheric prevailing wind and temperature profiles may significantly influence tropospheric ultralong wave propagation because the waves are partially reflected at internal boundaries. Figure 3 taken from Shutts [1978] shows as an example the height structure function of a thermally forced  $m = 1$  wave for two different prevailing wind profiles.

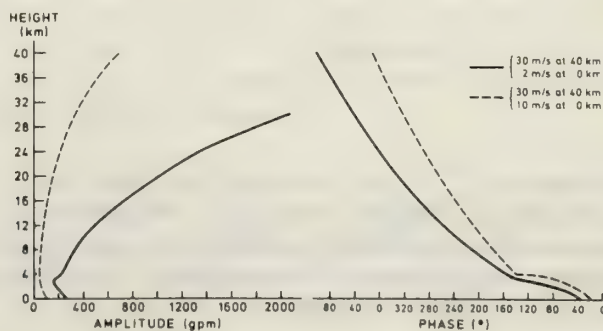


FIGURE 3: CALCULATED THERMALLY FORCED  $m = 1$  MODE IN TWO DIFFERENT CONSTANT ANGULAR SHEAR ZONAL WIND FLOWS. HEAT INPUT IS  $1.5 \text{ K day}^{-1}$  FOR  $z < 4 \text{ km}$  AND ZERO ABOVE [ADOPTED FROM SHUTTS, 1978].



Calculations based on slightly different concepts but with similar results have been carried out by Bates [1971], Geller and Avery [1977] and by Shutts [1978]. Whether such mechanisms may be responsible for long term variations in the climate, remains pure speculation, however, because our knowledge about long term variations in the stratospheric parameters is spurious and contradictory [see e.g., Pittcock, 1978]. Moreover, the forcing functions as well as the relative importance of thermal and orographic forcing are still not well explored [Saltzmann, 1968], and the theory of ultralong waves is far from being well understood [e.g., Hayashi and Golder, 1977].

#### 4. OBSERVED FEATURES OF PLANETARY WAVES

##### 4.1 Space-time Analysis at the 500 mbar Level

The theory of planetary wave propagation predicts eastward and westward travelling waves, mainly of the type of the internal class II waves, separated according to their zonal wave numbers  $m$  and their periods  $\tau$  (or frequencies  $\omega = 2\pi/\tau$ ). While the meridional structure of the wave mode is specified by the Doppler shifted frequency  $\omega_D$  (eq. 10) and therefore by the prevailing wind  $U_0$  and the frequency  $\omega$ , the direction of wave propagation is determined by the phase velocity  $c = a\omega/m$ , and therefore by the forcing function; e.g., a standing heat source  $Q$  with wave number  $m$  and frequency  $\omega$  is the generator of eastward and westward propagating waves because of the relationship

$$Q = Q_0 \cos m\lambda \cos \omega t = \frac{Q_0}{2} \{ \cos(m\lambda - \omega t) + \cos(m\lambda + \omega t) \} \quad (16)$$

It is appropriate, therefore, to separate the available observations according to wave number  $m$ , period  $\tau$ , and direction of propagation, by a space Fourier and time cross-spectral analysis [Hayashi, 1971]. Here the data for each latitude zone and for each day are first separated into modes of wave number  $m$  by a Fourier analysis. Thus, the higher order "noisy" modes are eliminated from the modes of lowest wave numbers (usually  $m < 10$ ). The new time series with daily values of sine ( $S_m$ ) and cosine ( $C_m$ ) terms for each mode and each latitude zone are used to compute power spectral estimates and cross amplitude/phase spectral estimates between  $S_m$  and  $C_m$ . From those, a decomposition into east and westward propagating waves and a standing wave is possible [Willson, 1975; Pratt, 1976; Schäfer, 1979].

Figure 4 shows the results of Schäfer [1979] who used the data of the 500 mbar level on the northern hemisphere for the years 1949 till 1976. The relative power of the first 6 modes of the travelling waves is plotted versus frequency for colatitudes between  $5^\circ$  and  $75^\circ$ . Westward travelling waves are represented by positive values, eastward travelling waves are



represented by negative values. In Figure 4, we notice two separate areas with eastward travelling waves, mainly at higher modes, and westward travelling waves, mainly at lower modes. The separating diagonal stretches approximately from  $75^{\circ}\text{N}$ ;  $m = 1$ , to  $20^{\circ}\text{N}$ ;  $m = 6$ . Two regions of large amplitude westward propagating waves can be distinguished for  $m = 1$ : one region at lower latitudes with largest amplitudes near periods of 5 days, and a second region at middle and higher latitudes with largest amplitudes near a period of 18 days. Eastward travelling waves for  $m = 1$  only exist at latitudes beyond  $75^{\circ}\text{N}$  with periods smaller than 7 days.

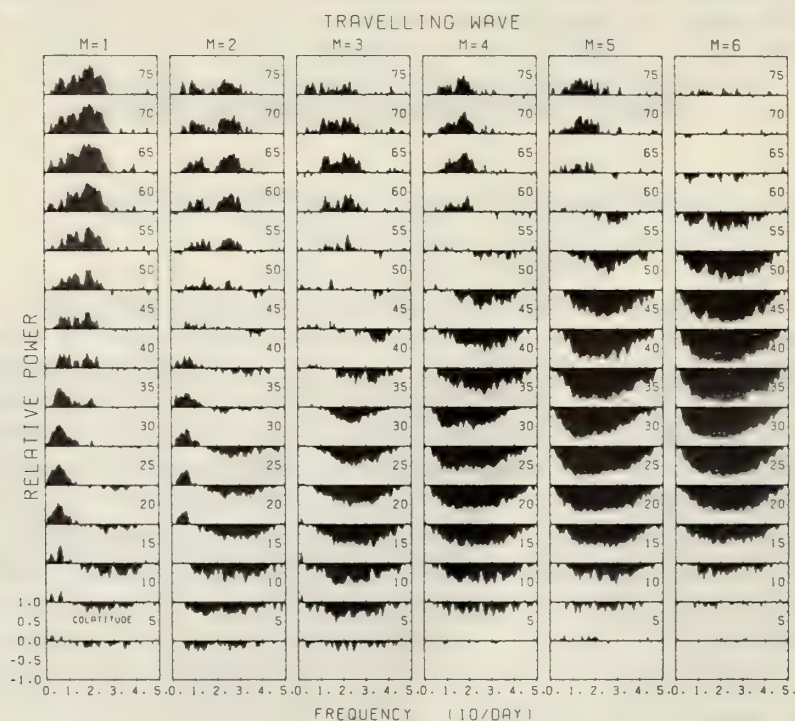


FIGURE 4: RELATIVE POWER SPECTRA OF TRAVELLING WAVES UP TO MODE  $m = 6$  VERSUS FREQUENCY AT COLATITUDES  $5^{\circ} - 75^{\circ}$ . WESTWARD TRAVELLING WAVES ARE DEFINED POSITIVE, EASTWARD TRAVELLING WAVES NEGATIVE.

These features can also be seen for  $m = 2$  waves where, however, the lower latitude maximum splits into two submaxima near 10 days and 4 days, and the zone of eastward travelling wave extends to  $60^{\circ}\text{N}$ . The standing wave structure, not shown here, has much smaller amplitudes and displays a noisy character. The westward propagating waves in the upper left from the diagonal in Figure 4 are the ultralong waves.

The preferred periods of the eastward propagating waves in the lower left from the diagonal in Figure 4 are between about 3 and 10 days. These eastward travelling waves can be interpreted as the synoptic scale Rossby-Haurwitz waves, observed at middle and higher latitudes.

The result of Figure 4 is consistent with results of Madden and Stokes [1975] and Fraederich and Böttger [1978] who used shorter time intervals and narrower latitude bands.

## 4.2 Origin of Westward and Eastward Propagation of the Waves

The preference of one direction of wave propagation in either type of waves may be qualitatively interpreted as follows: the synoptic scale waves propagate as free waves. The horizontal divergence of these waves is small compared to their vorticity. Therefore, their phase velocities are determined by the dispersion equation (12). For  $U_0 \approx 12$  m/s and  $6 < |n| < 9$ , it is  $-U_0 < c < 0$ . These waves are therefore predominantly westward travelling waves with  $|c| < U_0$ , and they appear to move to the east within the prevailing westerly wind as viewed by an observer on the earth. It is that relative eastward movement of the wave structure which the space-time analysis reveals.

On the other hand, the ultralong waves apparently are generated by thermal and/or orographic forcing [e.g., Saltzman, 1968]. Therefore, their direction of propagation is determined by the source terms. From reasons which are not yet known, propagation conditions for waves generated by the westward travelling source terms (e.g., the second term on the right hand side in (16)) appear to be better than the propagation conditions for waves generated by the eastward travelling source terms. Present theoretical models predict the contrary [Hayashi and Golder, 1977].

## 4.3 Quasi-Persistency of Planetary Waves and of Solar Active Longitudes

It can be shown that the waves in Figure 4 are quasi-persistent with life times of several periods. As an example, in Figure 5a the cosine terms of the Fourier spectrum of the  $m = 1$  wave at  $60^\circ\text{N}$  are displayed as daily deviations from a 27 day running mean. Each row in Figure 5a contains  $4 \times 27 = 108$  days of data, black areas indicating positive deviations, white areas indicating negative deviations. Each following row is shifted by 27 days, so that the second, third and fourth columns are merely repetitions of the first row. Figure 5b shows for comparison a random series with equal variance.

We notice from Figure 5a a recurrence tendency of waves with periods between about 25 and 30 days. Lower and higher periods are filtered out by the 27 running mean process. The waves in Figure 5a have life times from several months to about one year. Very probably, they are randomly generated by internal turbulent processes within the troposphere. The same is true for the other spectral bands seen in Figure 4.

For comparison, in Figure 6 the direction of the interplanetary magnetic field (left) and daily deviations from a 27 day running mean of the relative sunspot number (right) are plotted in the same manner as in Figure 5. In Figure 6a we notice the well known long living two and four sector structures with periods between 26 and 28 days [Svalgaard and Wilcox, 1975]. Likewise, the sunspots are grouped on preferred active longitudes which rotate nearly rigidly with the sun. They also have life times from months to years. However, no coherence exists between the sector structure of the IMF and the active longitudes for time intervals longer than the life times of the active longitudes.

The interplanetary magnetic sectors are believed to be magnetic manifestations of quasi-persistent slow hydromagnetic waves generated randomly within the turbulent convection zone of the sun [e.g., Suess, 1975]. The



active longitudes are related to the magnetic sectors in an as yet unknown way.

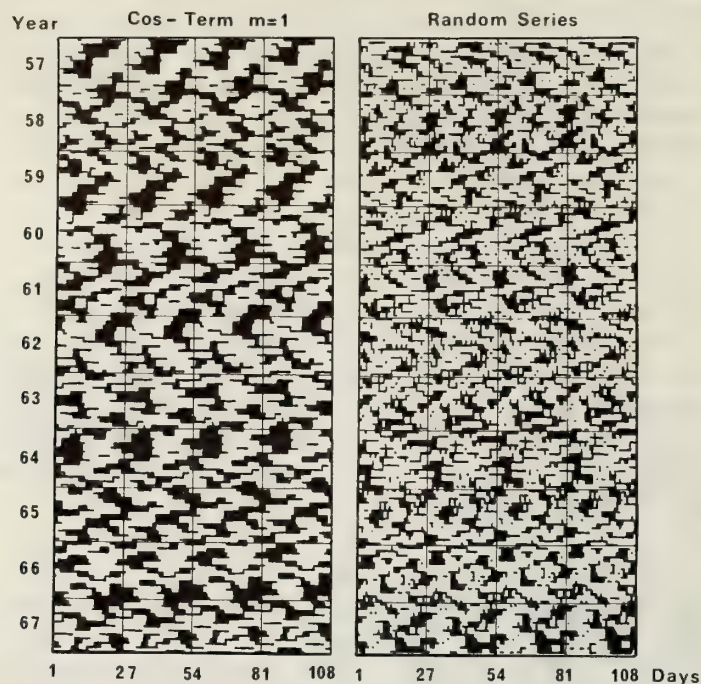


FIGURE 5: DAILY DEVIATION FROM A RUNNING 27 DAY MEAN OF THE COSINE TERM OF THE WAVE WITH ZONAL WAVE NUMBER  $m = 1$  AT  $60^{\circ}\text{N}$  LATITUDE AT THE 500 MBAR LEVEL FOR THE TIME INTERVAL 1957 TILL 1967 (LEFT), AND A RANDOM SERIES OF EQUAL VARIANCE (RIGHT). BLACK AREAS INDICATE POSITIVE DEVIATIONS, WHITE AREAS INDICATE NEGATIVE DEVIATIONS. EACH ROW IS DISPLACED BY 27 DAYS AGAINST THE PRECEEDING.



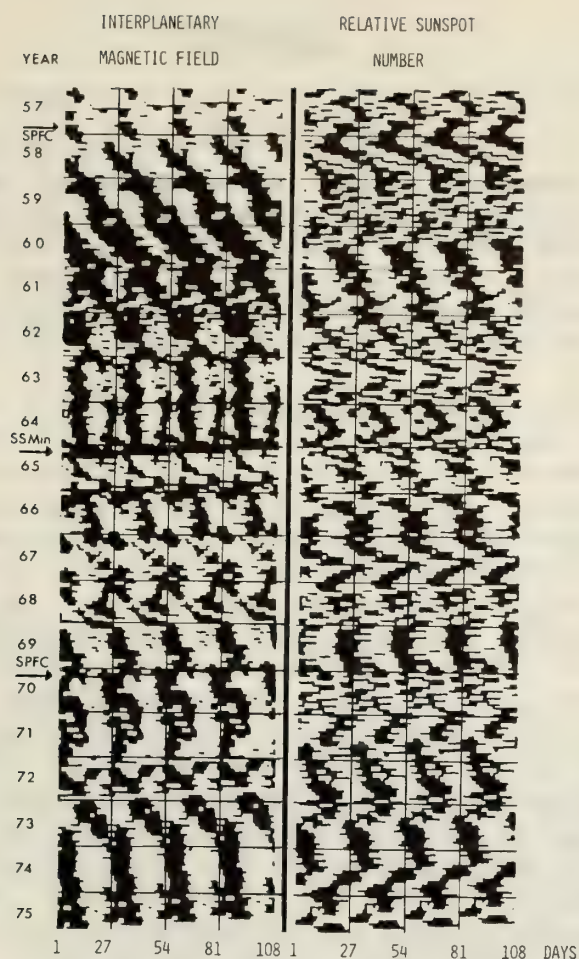


FIGURE 6: DIRECTION OF THE INTERPLANETARY MAGNETIC FIELD FOR THE TIME INTERVAL 1957 TILL 1975 [SVALGAARD AND WILCOX, 1975] {BLACK AREAS INDICATE AWAY SECTORS, WHITE AREAS INDICATE TOWARD SECTORS. EACH 27 DAY CYCLE IS REPEATED FOUR TIMES IN ONE ROW} (LEFT); AND DAILY DEVIATION FROM A 27 DAY RUNNING MEAN OF THE RELATIVE SUNSPOT NUMBER {BLACK AREAS INDICATE POSITIVE DEVIATIONS, WHITE AREAS INDICATE NEGATIVE DEVIATIONS} (RIGHT).

From a comparison between Figures 4 to 6, the following conclusion can be drawn. It is an accidental coincidence that in two completely independent systems turbulent processes generate quasi-persistent wave structures which both have similar preferred periods and comparable life times. If one observes both systems during a time interval which is not long compared with the life times of the turbulent cells and if one uses ordinary statistics of independent variables, the result gives the false impression of a real physical connection which in fact does not exist.

## 5. CONCLUSION

If sun-weather effects should have any meaning in weather and climate, they must manifest themselves in the form of planetary waves. From our present knowledge of the sun's short term radiative variability which is presumed to be  $\lesssim 1 \text{ W/m}^2$ , one can estimate that pressure waves at the ground of not more than about 0.1 mbar amplitude can reasonably be expected if active longitudes rotate rigidly with the sun, and that these amplitudes become even smaller if the time interval of solar activity is shorter (e.g., solar flares).

On the other hand, a space-time spectral analysis of planetary wave propagation reveals that broad spectral bands of quasi-persistent waves exist in the troposphere with relatively large amplitudes and life times of several periods, probably generated by internal turbulent processes.

Two types of waves can be identified: ultralong waves with low zonal wave numbers ( $m \lesssim 4$ ) travelling predominantly to the west and having periods between about 4 and 40 days with peaks near 5 days at lower latitudes and 18 days at middle and higher latitudes, and synoptic scale Rossby-Haurwitz waves with wave numbers  $m \gtrsim 3$  travelling predominantly to the east and having periods between about 3 and 10 days. The observed short term "sun-weather effects" probably are merely manifestations of these quasi-persistent waves which one is able to filter out of the meteorological data at almost any period including the period of solar rotation and its harmonics. The accidental coincidence of quasi-persistent turbulent cells on the sun which rotate nearly rigidly, and quasi-persistent planetary waves within the troposphere of similar life times and similar periods, can give the illusion of a physical connection between active longitudes on the sun, or sector structures of the IMF, with events in the troposphere.

The whole effort in "sun-weather" research may turn out to be not completely futile however. It may excite a stronger interest of the meteorologists in the study of the generation and propagation of ultralong waves and might ultimately improve long range weather forecasts. In this respect, it is perhaps worthwhile to remember that the astronomer Dr. Andrew E. Douglass, then director of the Steward Observatory of the University of Arizona at Tucson, discovered the tree ring dating method in 1913 during his effort to prove a relationship between rainfall and the eleven year cycle of the sunspots.

## Acknowledgements

H. Volland thanks Prof. P. Hays for his hospitality during his stay at the Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan.



## REFERENCES

- Bates, J. R. (1977): Dynamics of stationary ultra-long waves in middle latitudes. Quart. J. R. Met. Soc. 103:397.
- Bhatnagar, V. P., and T. Jakobsson (1978): Lack of effects of solar magnetic sector crossings on the troposphere. Geophys. Res. Lett., 5:180.
- Chapman, S., and R. S. Lindzen (1970): Atmospheric Tides, D. Reidel, Dordrecht, Netherlands.
- Charney, J. G., and P. G. Drazin (1961): Propagation of planetary-scale disturbances from the lower into the upper atmosphere. J. Geophys. Res., 66:83.
- Dickinson, R. E. (1973): Method of parameterization for infrared cooling between altitudes of 30 to 70 km. J. Geophys. Res., 78:4451.
- Foukal, P. V., P. E. Mack, and J. E. Vernazza (1977): The effect of sunspots and faculae on the solar constant. Astrophys. J., 215:952.
- Fraederich, K., and H. Böttger (1978): A wavenumber-frequency analysis of the 500 mb geopotential at 50°N. J. Atmos. Sci., 35:745.
- Geller, M. A., and S. K. Avery (1977): Calculations of solar activity effects on planetary wave propagation. Joint Symposium C. IAGA/IMAP Assembly, Seattle, Washington.
- Hayashi, Y. (1971): A generalized method of resolving disturbances into progressive and retrogressive waves by space Fourier and time cross-spectral analyses. J. Meteor. Soc. Japan, 49:125.
- Hayashi, Y., and D. G. Golder (1977): Space-time spectral analysis of mid-latitude disturbances appearing in a GFDL general circulation model. J. Atmos. Sci., 34:237.
- Hines, C. O., and I. Halevy (1977): On the reality and nature of a certain sun-weather correlation. J. Atmos. Sci., 34:382.
- Holton, J. R. (1975): The dynamic meteorology of the stratosphere and mesosphere. Meteor. Monogr., No. 37.
- Kasahara, A. (1976): Normal modes of ultralong waves in the atmosphere. Month. Wea. Rev., 104:669.
- King, J. W., A. J. Slater, A. D. Stevens, P. A. Smith, and D. M. Willis (1977): Large-amplitude standing planetary waves induced in the troposphere by the sun. J. Atmos. Terr. Phys., 39:1357.
- Larsen, M. F., and M. C. Kelly (1977): A study of an observed and forecasted meteorological index and its relation to the interplanetary magnetic



# REFERENCES (continued)

- field. Geophys. Res. Lett., 4:337.
- Livingston, W. (1978): Solar input to terrestrial system. Symposium on Solar-Terrestrial Influences on Weather and Climate, Columbus, Ohio.
- Longuet-Higgins, M. S. (1968): The eigenfunctions of Laplace's tidal equation over a sphere. Phil. Trans. Roy. Soc., London, 262:511.
- Madden, R., and J. Stokes (1975): Evidence of global scale 5-day waves in the 73 year pressure record. J. Atmos. Sci., 32:831.
- Markson, R. (1978): Nature., 273:103.
- Phillips, N. A. (1963): Geostrophic motion. Rev. of Geophys., 1:123.
- Pittock, A. B. (1978): A critical look at long-term sun-weather relationships. Rev. Geophys. Space Phys., 16:400.
- Pratt, R. W. (1976): The interpretation of space-time spectral quantities. J. Atmos. Sci., 33:1060.
- Saltzman, B. (1968): Surface boundary effects on the general circulation and macroclimate: A review of the theory of the quasi-stationary perturbations in the atmosphere. Meteor. Monogr., No. 30.
- Schäfer, J. (1979): A space-time analysis of tropospheric planetary waves in the northern hemisphere, J. Atmos. Sci., 36:1117.
- Shutts, G. J. (1978): Quasi-geostrophic planetary wave forcing. Quart. J. R. Met. Soc., 104:331.
- Suess, S. (1975): Hydromagnetic theory of solar sectors: slow hydromagnetic waves. AIAA Journal, 13:443.
- Svalgaard, L., and T. M. Wilcox (1975): Long term evolution of solar sector structure. Solar Phys., 41:461.
- vanLoon, H., R. L. Jenne, and K. Labitzke (1973): Zonal harmonic standing waves. J. Geophys. Res., 78:4463.
- Volland, H. (1977): Can sunspots influence our weather? Nature, 269:400.
- Volland, H., and H. G. Mayr (1977): Theoretical aspects of tidal and planetary wave propagation at thermospheric heights. Rev. Geophys. Space Phys., 15:237.

#### REFERENCES (concluded)

- Wallace, J. M. (1973): General circulation of the tropical lower stratosphere. Rev. Geophys. Space Phys., 11:191.
- Wiin-Nielsen, A. (1964): Further studies of energy exchange between the zonal flow and the eddies. Tellus, 16:168.
- Wiin-Nielsen, A., J. A. Brown, and M. Drake (1963): On atmospheric energy conversions between the zonal flow and the eddies. Tellus, 15:261.
- Wilcox, J. M. (1975): Solar activity and the weather. J. Atmos. Terr. Phys., 37:237.
- Wilcox, J. M., P. H. Scherrer, L. Svalgaard, W. O. Robert, R. H. Olson, and R. L. Jenne (1974): Influence of solar magnetic sector structure on terrestrial atmospheric vorticity. J. Atmos. Sci., 31:581.
- Willis, D. M. (1976): The energetics of sun-weather relationship: Magnetospheric processes. J. Atmos. Terr. Phys., 38:685.
- Willson, M. A. G. (1975): A wavenumber-frequency analysis of large-scale tropospheric motions in the extratropical Northern Hemisphere. J. Atmos. Sci., 32:478.

## THE CLOUDED CRYSTAL BALL: COMMENTS ON GEOPHYSICAL PREDICTION

S. M. Silverman  
Air Force Geophysics Laboratory  
Bedford, MA 01731

A prophet will the Lord thy God raise up unto thee, from the midst of thee, of thy brethren, like unto me; unto him ye shall hearken.... And if thou say in thy heart "How shall we know the word which the Lord hath not spoken?" When a prophet speaketh in the name of the Lord, if the thing follow not, nor come to pass, that is the thing which the Lord hath not spoken; the prophet hath spoken it presumptuously, thou shalt not be afraid of him.

Deuteronomy, 18:15, 21-22

There shall not be found among you any one that maketh his son or daughter to pass through the fire, one that useth divination, a soothsayer, or an enchanter, or a sorcerer, or a charmer, or one that consulteth a ghost or a familiar spirit, or a necromancer. For whosoever doeth these things is an abomination unto the Lord.

Deuteronomy, 18:10-12

### 1. INTRODUCTION: The Function of Prediction

This paper is concerned with some of the concepts of prediction in the geophysical domain. It is an attempt to clarify some of the areas in which difficulties arise and try to indicate the nature of these difficulties. These problems are approached from the viewpoint of many years as researchers in the field. We avoid academic exercises and discussions, preferring instead the viewpoint of the working scientist. Prophecy, the prediction of things to come, has a long history, probably as old as civilization itself. The practical utility of knowing the right time to plant crops, of unseasonable weather coming, of tornadoes, as examples is obvious. The usefulness of knowing the dangers that may befall an individual is equally obvious. Here, however, we enter into a new sphere: The capacity of free will to alter coming events. Prediction, then, from the earliest times until the present, has had at least two aspects, the foreordained occurrence of natural events, and the capacity to alter the consequences of these events or possibly the events themselves. This psychological dichotomy, between those who accept nature, and those who strive to alter it, may be at least one factor in the recurring battles among scientists dealing with predictions.

This conflict can have serious consequences for the individual. The theory of lunar eclipses, for example, was first written down by Anaxagoras, but was kept secret by an elite group. Plutarch notes: "People would not then tolerate natural philosophers, and theorists, as they then called them, about things above; as lessening the divine power, by explaining away its



agency into the operation of irrational causes and senseless forces acting by necessity, without anything of Providence or a free agent. Hence it was that Protagoras was banished, and Anaxagoras cast in prison.... "It may come as a shock to contemporary scientists to hear the bases of their efforts described as "irrational causes and senseless forces acting by necessity". The mechanistic view of science, popularized in the early nineteenth century, is equivalent to a conception that everything is foreordained - a concept which we usually think of as being predominant in the mystical east. With the western scientists' belief in quantum mechanics, the probability of free will appears very tenable. Thus the ethos of contemporary western man has been that everything can be accomplished if we work hard enough at it. More recently technology, and this view, have come under closer scrutiny. In other words forces greater than ourselves sometimes control our lives. This can lead to a type of mental impotence, whereby the mental energies of society are not used to resolve its problems. Thus science is viewed as a tool to help society, but with its awesome power, we must always consider "safety comes first!"

The growth of science in the past centuries has led to a new function for prediction, that is, as a test of hypotheses. It is perhaps the cornerstone of the scientific method that the validity of a theory or hypothesis is tested by predicting a result which is found to be correct either observationally or experimentally. This approach raises some new problems. For example, is a theory that predicts a result unique or not? Does the fact that a prediction is correct prove the validity of a theory or is it possible that one or more other theories would give the same result? If so, are these various theories all equivalent to or derivable from one another?

The functions of prediction are thus at least three-fold in nature. There is, first of all, the prediction of future events for planning purposes. The time of annual flooding of the Nile is needed in order to know when to plant. Departmental budget projections are needed in order to know whether an additional person can be hired. The motivation for funding of sun-weather research seems to depend largely on this eventual utilitarian product.

Secondly, we need to know the future in order to take action to avert catastrophes. In distinction from the planning function here we must know the likelihood of individual events. Statistical results are not of much use to us. For some the avoidance of catastrophes (as in popular belief in the Anaxagoras case above) lies in placating the gods, or in repentance. For the secularist the knowledge of specific impending events simply allows mitigating actions to be taken. The interest in earthquake predictions in recent years is to allow for the evacuation of residents in threatened areas. Forecasts of the path of a hurricane allow people to retreat to safe places and to secure their property as well as possible.

Finally, there is the function of hypothesis testing. Here the focus is simply on whether the predicted event occurs. Whether the light of a star is bent by the gravitational field of the sun becomes a test of relativity theory. The required prediction may be of an already known observational fact. A dynamic theory of sunspots must predict an average eleven year cycle. If additional cycles are verified then the theory must predict these as well. If the theory predicts one observation but then fails to predict another, then the theory must be modified or replaced. Looked at from this point of view the crux of the test becomes also, how or when do we

know when a prediction has been verified? Before returning to this question we shall take a short detour into belief.

## 2. BELIEF: Public and Private

It is quite clear that we all know more than we can say or articulate. To take a simple example, consider the act of moving your hand. All of us know how to do this, but it is unlikely that there is anyone who can state or explain in detail how this is done. There is first of all the question of mind and will: What are the processes of thought, what converts a thought (whatever that is) into an impulse travelling along some pathway. What is the nature of this pathway, the mechanism of transmission of the impulse, and how it gets to wherever it is going. Finally, what are the nature and mechanism of the motor action that results in the movement of the hand. It is clear that we are constantly acting without knowing how (some people say even without knowing why). C.S. Peirce, the scientist best known as a philosopher and the inventor of pragmatism, pointed out (Collected Papers, 1958, 7:36-46) that in our daily life we are constantly making judgments and decisions on the basis of incomplete information, and that this is a necessary part of life. He asks, what would happen to the bird who refused to leave the nest until he thoroughly understood all the principles of aerodynamics? Some of you who know perfectionist scientists who refuse to publish until every last detail is thoroughly and completely worked out may infer their motivation from this comment.

We may then recognize a distinction between what we may call reason, and the unarticulated thinking process we may call instinct. Peirce (Collected Papers, 1932, 1:616-677) notes that in our daily life we must, and do, act on instinct, which is more basic: "It is the instincts, the sentiments, that make the substance of the soul. Cognition is only its surface, its locus of contact with what is external to it." (1:628). Also (1:631):

Reason is of its very essence egotistical. In many matters it acts the fly on the wheel. Do not doubt that the bee thinks it has a good reason for making the end of its cell as it does. But I should be very much surprised to learn that its reason has solved that problem of isoperimetry that its instinct has solved. Men many times fancy that they act from reason when, in point of fact, the reasons they attribute to themselves are nothing but excuses which unconscious instinct invents to satisfy the teasing "whys" of the ego. The extent of this self-delusion is such as to render philosophical rationalism a farce.

This tension between the conscious process of reason, and the unconscious process of instinct, has a strong influence on what the individual scientist is willing to accept and on what he is willing to work on. This was noted, for example, in a study of auroral audibility by Silverman and Tuan (1973, at pages 158-159).

This tension leads to the distinction between what we are calling here public belief and private belief. Private belief is what you know to be



true or correct. Public belief is what you are willing to state in a publication or a lecture or, in general, to your fellow scientists, other than those you know very well. Our private belief sometimes appears in print disguised by such phrases as "the data indicate that...." or "it is felt that ...." or some such similar wording. This allows for the instinctual knowledge developed from long familiarity with the data to be put forward, albeit cautiously. Public belief, on the other hand, depends on the consensus by the community of scientists of what constitutes adequate proof. This, in turn, depends on what we have been taught is correct.

Generally, among physical scientists, quantitative predictions quantitatively verified are acceptable. The physical scientist looks with disdain on the soft, often multi-valued domain of the social sciences.

Freudian psychoanalysis, for example, makes the physical scientist extremely uncomfortable. But even in the physical sciences we run into problems. In the determination of a spectrum, for example, there is always a point at which the line is only slightly greater than the noise, and may, in fact appear on some measurements and not on others. Can one state publicly that the line is there, even though all of one's experience says that it is real? Of course, higher resolution techniques and better signal processing techniques are helpful but there will always be a point at which the signal is comparable to noise. In questioning several experimentalists on this point some years ago, a common answer was that if the signal was about 10% greater than the noise it was believable. Examples of this sort could be multiplied by any experienced experimenter or theoretician.

We should recognize that the use of a consensus leads to a lowest common denominator of proof. It is, after all, what everyone can agree on. Beyond that there is always room for argument. Quantitative mathematical techniques are not immune from these problems. When does one believe statistical data, for example? Statistical measures are generally (but not always) based on a normal probability distribution for the parameters. How often does the experimentalist demonstrate that this is true for the quantities he is measuring? Have sampling fluctuations been taken into account, and how does one do this? The examples could easily be multiplied.

The problem is made considerably worse when there is an absence of direct, one-to-one causation. This is certainly true for sun-weather relationships, and we suspect that much of the problem in this field arises from differences between those who know certain things to be true for this complex fabric, but can only partially demonstrate them, and those who insist on the rigorous quantitative proofs of consensus science.

#### CAUSATION: The Chain of Causality

We may illustrate a problem of causality with a case from the legal literature, Chase vs Washington Water Power Co., 111 P. 2d 872 (Idaho 1941). The plaintiff sued the power company for damages caused by a fire in his barn, for which he attributed liability to the power company. The company had a power line running along the highway. The high voltage line was attached to an insulator attached to a pole. The pole had a guy wire running to the ground and separated from the high tension wire by a standard distance of 28 inches. Along the highway a three strand barbed wire fence ran, enclosing the land of the plaintiff. At one point this fence was attached to the plaintiff's barn. Over a period preceding the date of the event the fence had leaned over so far out of line that its three strands of wire touched and rubbed against the uninsulated guy wire. This



sets the scene. On August 25, 1939 two chicken hawks were engaged in an aerial battle. Their talons were interlocked. In the course of the battle (this is the theory accepted by all parties) one of them touched the high-tension transmission line at the same instant that the other touched the guy wire, thus forming a circuit between the two. As a result numerous fires were observed in the dry grass, all along the barbed wire fences, and at the barn. The barn was totally destroyed and some other buildings were injured by the fire. We have here a relatively simple chain of causation. Consider, however, some of the various things which could have happened at different parts of the chain which could have produced a different result. The grass, and thus presumably the barn, was dry in an August heat. What if it had been properly maintained? And the most easily perturbed part of the chain is the chicken hawks aerial battle. We do not pretend to be familiar with the psychology of chicken hawks, but there must be many factors which could have led either to no battle or to a battle in a different location. Also air currents, atmospheric structures, whatever affects the aerodynamics of chicken hawks would play a role in determining their presence at that moment in that particular location. Variations in a great many parameters would have produced a different result. The fact of causation of a barn fire is indisputable. It is unlikely, however, that any theory, accepted as reasonable, by the scientific community, would have predicted this result. Prediction would depend on a strong deterministic theory of the early eighteenth century variety. Nor would statistical data have shown that barn fires would be caused by the irritations of chicken hawks. A correlative study of the emotions of chicken hawks and barn fires in Idaho would almost certainly show a correlation coefficient close to zero. This simple example indicates that when the chain between input and output has several links in it we must make a sharp distinction between individual events and statistical correlations.

The sun-weather system, compared to the above, is a veritable chamber of horrors. Consider the accompanying figure (from Kellogg and Schneider, 1974, reproduced in J.R. Herman and R.A. Goldberg, 1978, P.5). Here we do not have a simple chain of causation, we have a complex web of inter-relationships. It is surprising that any sense can be made out of it at all. Variations in a single parameter may have effects in areas not considered because they have not been included in the web. Consider, for example, the production of energy from carbon based fossil fuels. Atmospheric optical properties are changed, the temperature of the surface may increase, perhaps other mechanisms may come into play, such as changes in global plant growth. What is the overall net result, and in view of the many intervening processes and mechanisms, can we predict this result, and to what accuracy?

Let us return for a moment to the rigorous world of the celestial mechanician. Here, with Newton's Laws, we can predict with great accuracy the ballistic trajectories of space craft, eclipses of the sun and moon and similar phenomena. Our predictions have a high degree of validity, at least over the short term. We are not surprised that when an eclipse is predicted for a particular location at a particular time it actually takes place according to our predictions. We are only surprised when the reality differs from our predictions, and we infer that some modification of the

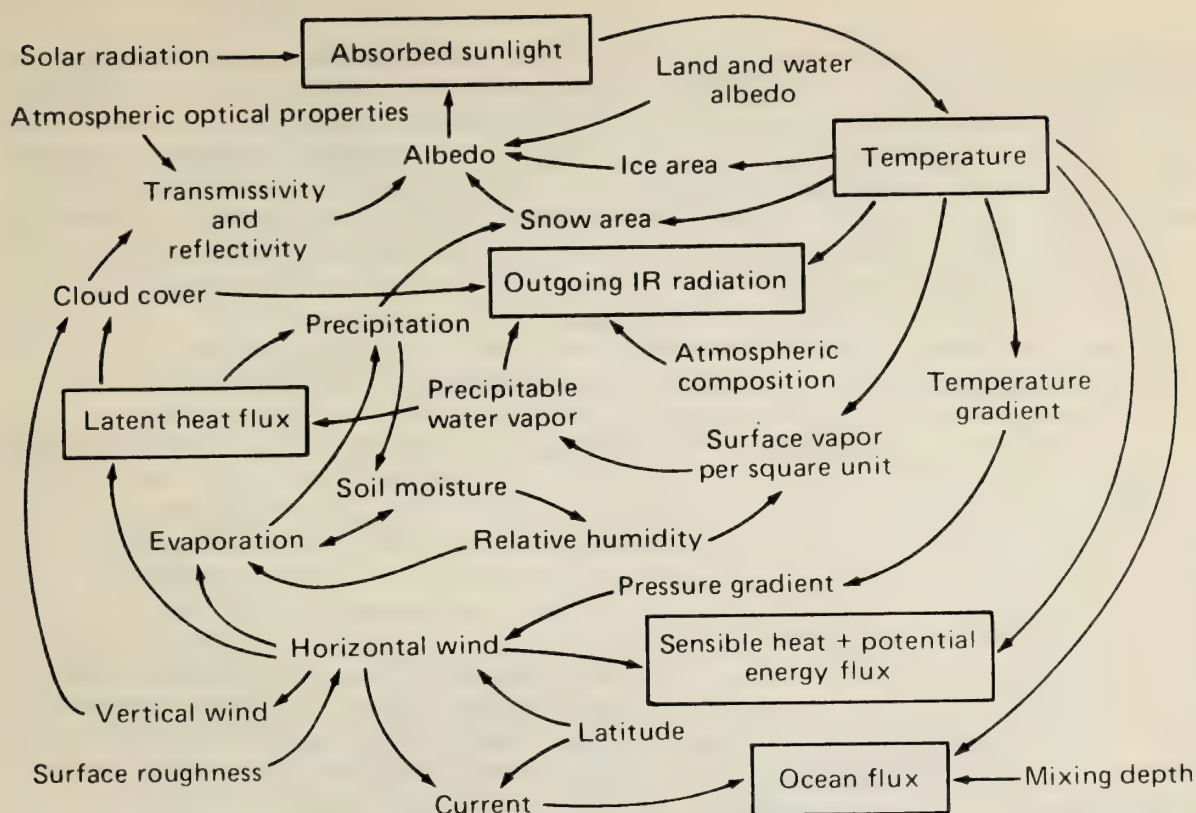


FIGURE 1

theory is needed. Tidal effects, varying over time, for example, will affect our eclipse predictions. The quality of our predictions is dependent on the number of unproven ad hoc assumptions we must make, and the predictions serve as a test of these assumptions. But, basically, we arrange, to the maximum extent possible, our observations in such a way that there is only a single step between cause and effect, and only a single cause which is varied. This procedure becomes difficult, if not impossible, in the real world of solar-terrestrial predictions.

#### CAUSATION: Likely and Unlikely Events

In order to predict we must find "cause" and the quantitative "laws" which relate these causes to the effects we are looking at. We are, then, looking at the relationship between two observed parameters. Since the parameters we are interested in are rarely well behaved in the real world the study of an individual event may guide but not determine our conclusions. Carrington (1859), in the observation which initiated contemporary solar-terrestrial relationships studies, noted that "one swallow does not a summer make". We therefore tend to look for statistical relationships. These are tests from hindsight. Most commonly they take the form of correlations, superposed epoch analyses, power spectra, Fourier analyses, and similar methods. These techniques, however, can be, and are, criticized as being valid only if there is some physical mechanism underlying the particular correlation in a cause-effect manner. Thus, for example, there is a well known high correlation in the United States between the number of school teachers and the consumption of alcohol. Are we to infer from this



that education leads to alcoholism? The correct explanation, of course, is that the population of the United States has been increasing, and as it does both the number of school teachers and alcohol consumption increase simply because the population from which both are drawn is increasing. Thus statistical relationships by themselves are insufficient. One must have theory, perhaps suggested by the correlations found, which predicts these correlations. To the extent that our chains of causation are long with the intervening links each susceptible to modification by many factors, theory making can be a difficult and unrewarding task.

A further complication arises when we consider individual events, which may be the product of unlikely causes. From the practical point of view, however, these individual events, no matter how unlikely, are extremely important. Consider, for example, the consequences of an unlikely earthquake with epicenter at the site of a nuclear reactor. For these events our modes of prediction are likely to be quite different than statistical procedures. We can illustrate these freak situations with two examples drawn from the legal literature. In the first case, United Novelty Co. vs. Daniels, 42 So. 2d 395 (Miss. 1949), an employee was cleaning coin-operated machines with gasoline in a room eight feet by ten feet in an area which was heated by a gas heater with an open flame. The room thus contained gasoline vapors. The immediate cause of the explosion which fatally burned the employee, however, was a rat which had been in the machine and had thereby had its coat impregnated with gasoline. The rat escaped from the machine, sought sanctuary under the heater, caught on fire and retreated to its original hideout in the machine, thus setting off the blast. In the second case, Ramsey vs. Carolina-Tennessee Power Co., 143 S.E. 861 (N. Carolina 1928), a workman was killed by the shunting of a railroad car several miles away. The railroad had a spur track onto which they occasionally shunted cars. The track had no bumper or embankment at the end of it. Past the end of the track, but on the right of way of the railroad, a power company had erected a pole with a guy wire. On the pole was a transformer which converted the high voltage into a lower voltage which went into town by several lines. One of these led to a steam laundry which used washers operated by electricity. On the day of the accident the railroad applied a little too much force in shunting cars into the spur track. One car ran off the track into the field, hit the power company pole, damaging the transformer, and transmitting the high voltage current along the line to the laundry. The employee, in attempting to switch off the current to the smoking washing was electrocuted.

In hindsight each of these can be understood if we are in possession of all of the facts. They are only partially useful for predictive purposes. We can say, on analysis, that the situating of a power line pole not too far from the end of a track with no stop and which is used for shunting cars is an invitation to an accident of some kind. The accident may not occur at all. If it does it may scratch the pole without further damage, may seriously damage the pole without affecting the transformers, the workmen may or may not attempt to switch the washer off, and numerous other possibilities. We are left with many options, but are unclear as to which will actually occur, if any. Even if we know that an earthquake is likely to occur in a particular locality, can we predict which buildings will be damaged, which people will be in them, or will these depend on the time of day, the condition of the ground, building excavation along the



path of the shock, or many other factors? If there is a flare on the sun, to what extent can we predict the terrestrial effects? This will depend, in part, on the heliographic latitude and longitude, on the condition of the interplanetary medium at the time, and many other factors, depending on the effect we are asking about. Under these conditions what are our criteria for prediction? How rigorous can we be in testing these predictions? We are not suggesting that it is impossible to either predict or test predictions, merely that we must be more careful in defining the bounds of predictability and of the validity of predictions.

#### THE PERILS OF EXTRAPOLATION

We have in the previous section discussed essentially some problems of hindsight. What happens when we use the results obtained by hindsight to try to predict the future? If we have a well behaved system, as the revolutions of the planets about the sun, we can do pretty well even without a theory. But what if we have a series of empirical results with a multi-linked chain of causation, and without well-established underlying mechanisms for each step in the causal chain? What happens can be illustrated by those few intrepid researchers who have had sufficient faith in their harmonic analyses of sunspot occurrence to extrapolate them outside the range of the data. In Figure 2 we show the results obtained by Kimura (1913) and by Dinsmore Alter (1928). In Table I we have compared the predicted epoch of maximum and the sunspot number at maximum with the observed. In general Kimura's predictions of the maximum epoch are reasonably close to the actual, indicating some validity for the harmonic analysis. Alter's, past the first two, seem widely in error. For the actual amplitude of the curve at sunspot maximum both predictions are, overall, sufficiently seriously in error so that we could not trust them for predictive purposes. A similar result is obtained from a comparison of predictions which have been made for the current cycle, shown in Table II taken from H.H. Sargent, III, "A prediction for the next solar cycle" (1978) (unpublished). It is clear that the 16 predictions ranging from 48 to 200 for the maximum of the current cycle, most must be wrong. Judging by recent activity a number of these predictions have already proven wrong.

It is clear that foresight alone is insufficient. We may hope that if we can develop a sufficient understanding of underlying physical mechanisms then our foresight may improve. The result may be, however, that we may have to accept a measure of randomness in our theory, as in a Markov process, and our predictive capacity may be correspondingly limited. A recent paper on sunspot cycles, in fact, (J.A. Barnes, H.H. Sargent, III, and P.V. Tryon, "Sunspot Cycle Simulation using random noise", unpublished) has treated the sun as a resonant system with random impulses and has resulted in calculated cycles similar in many respects to those observed.

#### HYPOTHESIS TESTING AND INVERSE PROBABILITY

If we can describe a situation exactly, we can forecast the probabilities resulting from that situation. But what if we know the result? Can we then work backwards to deduce the original situation? In deductive science we typically assume a model, deduce a particular functional form relating different quantities from this model, and then see whether this is experimentally verified. If this is found to be the case then the argument is that the result tends to confirm the correctness of the assumed model.

FIGURE II ANALYSIS AND EXTRAPOLATION OF SUNSPOT NUMBERS

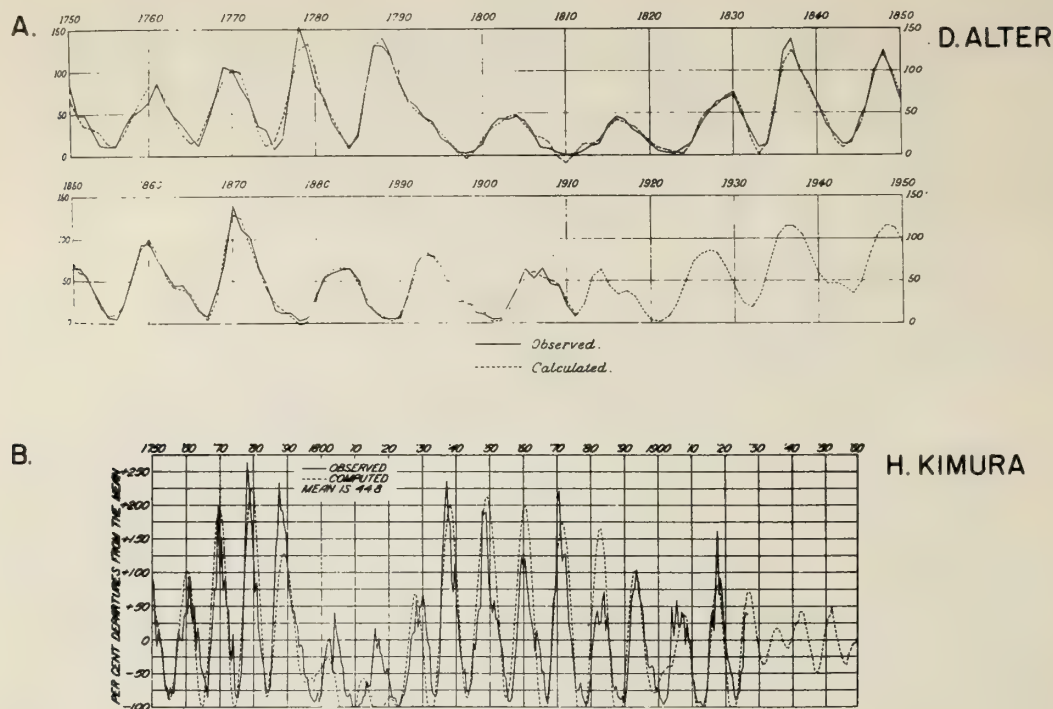


TABLE 1 COMPARISON OF PREDICTED AND OBSERVED SUNSPOT NUMBER MAXIMUM AND EPOCH OF SOLAR MAXIMUM

A. Comparison with results of Kimura

$\Delta$  = difference between actual and observed

Predicted Epoch Year	Sunspot Maximum	Actual Epoch Year	$\Delta$	Sunspot Maximum	$\Delta$
1914	65	1917	+3	104	+39
1927	85	1928	+1	78	- 7
1936.5	115	1937	+0.5	114	- 1
1948	115	1947	-1	152	+37

B. Comparison with results of Alter

1927	77	1928	+1	78	+ 1
1935	53	1937	+2	114	+61
1942	64	1947		152	+87
1951	67	1957		190	

TABLE II

A listing of predictions for Cycle 21 which has been published during the past two solar cycles.

Authors	Predicted Maxima	Epoch
Bonov	50 (+10)	1984
Cohen and Jacobs	48 to 96	1982
Cohen and Lintz	50	1982
Cole	60	1981
Gleissberg	56 to 96	1980
Henkel	100	
Hill	130 to 200	1981
King-Hele	110	1978
McNish and Lincoln	117 (see note)	1980
Schove	145	1984
Xanthakis	85	
Jose	low	1981
Smith	150	1980
Wood	60	1982
Ohl	140 to 180	1979

Note: Predicted value is taken from the October 1977 issue of the Solar-Geophysical Data Bulletin.

What we know experimentally is that a particular functional form can be used to treat the data. What we do not know, and can not know, is how many models, out of the universe of all possible models, will result in this particular functional form. We are therefore forced, in essence, to rely on a concept of proof which depends on inverse probability, that is, to what extent is the hypothesis proven by the result. This question has been discussed since the discovery of the theorem of inverse probability by Bayes in the eighteenth century and strong proponents can be found on both sides of the question, with no generally accepted solution as yet. The question is particularly troublesome when we are dealing with a general property of the system, such as the energy or the entropy, which change with every change in the state of the system. It seems reasonable that if a particular functional form is predicted and found, then this enhances the probability, since many other functional forms are possible, but the problem of quantifying the probability remains, since we do not know how many hypotheses can lead to even this particular form. The existence of analogies may perhaps make the hypothesis more plausible but does not necessarily increase the probability, unless we have independent reason to believe that corresponding quantities behave similarly.

An intriguing application of these concepts to astrophysical hypotheses was carried out a few years ago by P.A. Sturrock. He points out that in testing hypotheses one rarely has either as many observational data or as detailed or reliable theories as one would like. Furthermore he points



out the subjective nature of research, that each research worker decides how much weight to give to the statements of his colleagues. He then proposes a research formalism, a bookkeeping procedure, to evaluate hypotheses which take account of this subjective attribution of probabilities to both facts and hypotheses. The theorems of inverse probability are used for this purpose. We will not go through the details here because of lack of space. Sturrock applies the method as an example to the known result of the pulsar being a neutron star rather than a white dwarf. As far as we are aware his method has not been otherwise applied to real world situations. His example raises the question of whether a calculation of probabilities adds anything when sufficient information is available for the jury to come in with a verdict. Nevertheless the technique could conceivably be more useful for the more difficult and fuzzy areas encountered in sun-weather relationships.

#### DISCUSSION AND CONCLUSIONS

We have tried in this paper to identify some of the areas which cause difficulties in discussions of geophysical relationships. Our focus has been on the differences in defining and determining the validity and significance of hypotheses and observational correlations. This difficulty seems to arise primarily from the lack of direct one-to-one cause-effect systems. We are faced not simply with causal chains but with causal webs. Each link of the web is itself influenced by other factors, so that it may sometimes be operative in the chain and sometimes not. Thus correlative studies may indicate important factors but cannot definitively establish them, except in rare instances. The underlying mechanisms are also generally not precisely known or clearly defined as to their boundaries. Our knowledge of the entire system is thus primitive at best.

There are, furthermore, differences in psychological emphasis among different people. If we are not aware of these our discussions can easily degenerate into rigid stand offs between those who know intuitively and those who believe only what can be rigorously established. The quotations with which we have begun this paper are interesting in this respect. The question at issue was how to distinguish the true prophet from the false prophet. The answer given is simply to see whether the predictions come to pass. In its simplest essence this is what we are doing. But this is not all that is involved. Prophecy in Biblical times and before was clearly a well developed profession. The second quotation indicates that aids to prophecy were widely used. This was also true in the Roman and Greek world. In the Mosaic system truth did not come with the aid of instruments, but through internal revelation, what we have called intuition. Contemporary studies of creativeness have shown that the period of conscious preparation is succeeded by moments of creative insights, sometimes at much later times. The laws by which this happens remain a mystery still. The Biblical concept can thus be translated into contemporary terms by using instruments to collect data and to provide the material upon which the creative intuition operates to find truth. This truth must be verified by the success of its predictive capacity. Thus in our view tolerance is called for: we must have data collectors and correlators, but we must also have the speculators and theorists. These latter, however are obligated to provide predictions which are sufficiently detailed to be tested and verified, so that we may separate the true from the false prophets.

There still remains the question of agreement on what constitutes verification. Biblical prophecy concerned itself with tangible discrete events: the outcome of a battle, the death of a king. We are not so fortunate. We find fuzziness in our results, unlikely happenings. If we find correlations, what constitutes a significant correlation coefficient? Do we have a sufficiently known probability distribution so that we can determine significance? If we determine the power spectrum of a parameter, then how much above the continuum (red or white, and how determined?) does a peak have to be before we believe that it is significant? What are the tests for significance for a peak in a superposed epoch analysis? If we hypothesize a mechanism for sunspot production, do we insist that it predict one period of multiple periods before we begin to believe it? All of these are examples of difficulties in this area.

We feel that we must be less monolithic in our standards. We need to define different standards for direct causation and for causal chains or webs. We need to define the boundaries of significance tests, statistical or otherwise. And finally we need to accept the different roles, boundaries and standards of data gathering, data analysis and theory making including speculation. Only by being aware of these distinctions can we move forward to problem solutions without wasting our energies on useless squabbles amongst ourselves.

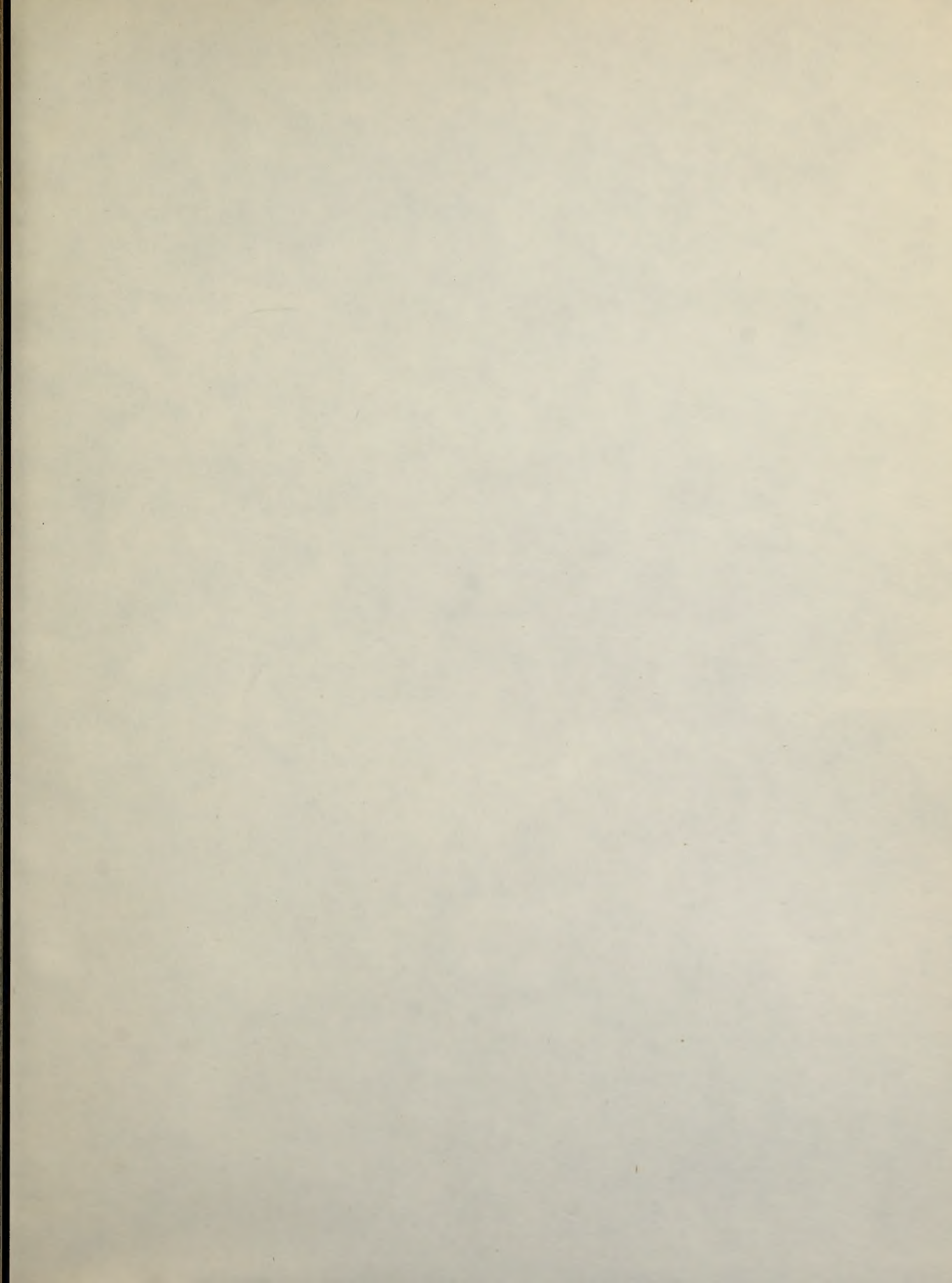
ACKNOWLEDGEMENTS: This paper was stimulated by discussions with Kenneth Schatten. I am indebted to him also for the opportunity to present it at the sessions of Working Group D at the Workshop.

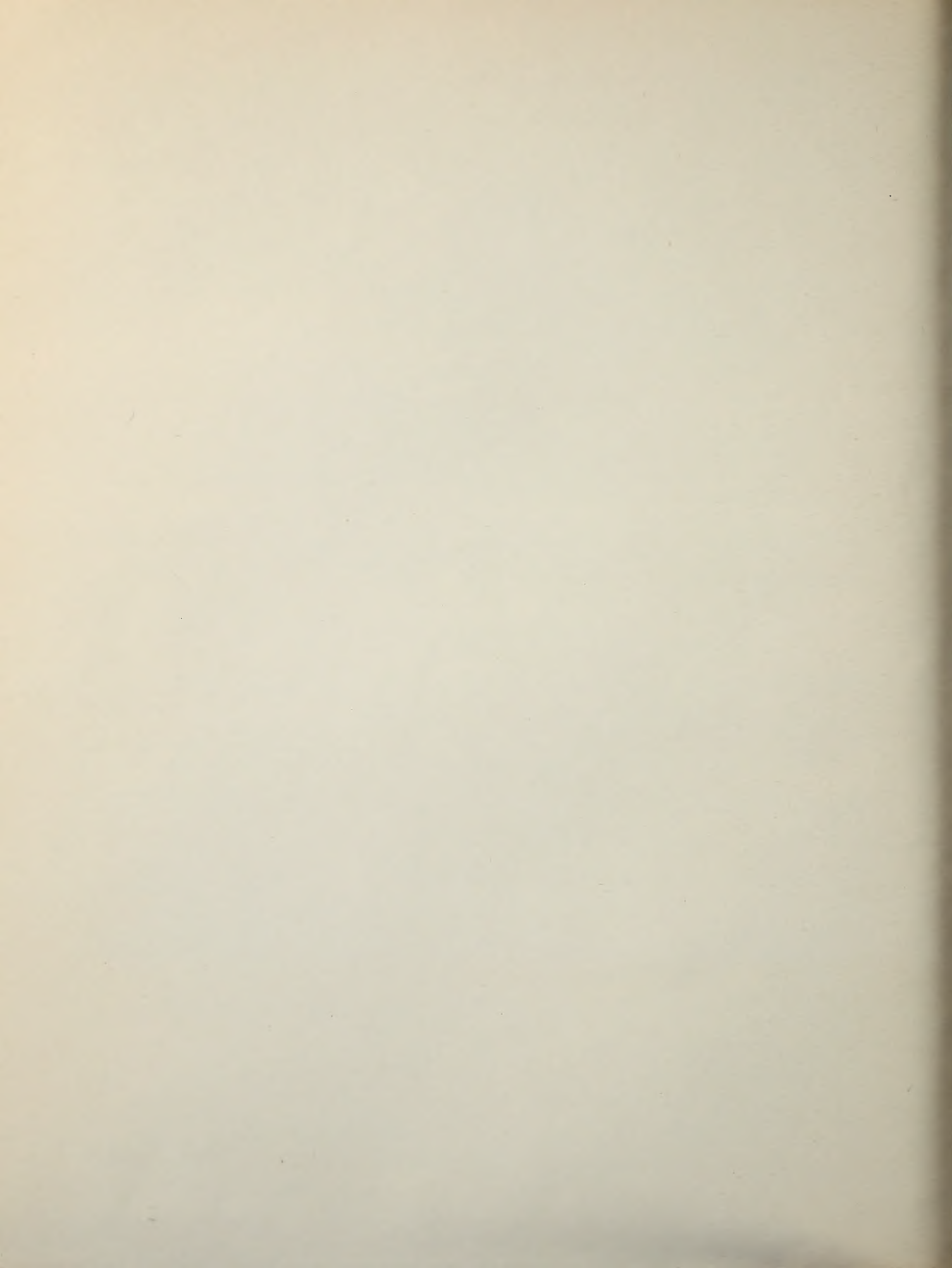
#### REFERENCES

- D. Alter, Monthly Weather Review, 56 399 (1928)
- R.C. Carrington, Monthly Not. Roy. Astr. Soc. 20 13 (1859)
- J.R. Herman and R.A. Goldberg, Sun, Weather and Climate, NASA SP-426 (1978)
- H. Kimura, Monthly Not. Roy. Astr. Soc. 73 543 (1913)
- C.S. Peirce, Collected Papers, Vol. I, Cambridge: Harvard Univ. Press, (1932)
- C.S. Peirce, Collected Papers, Vol. III, Cambridge: Harvard Univ. Press, (1958)
- S.M. Silverman and T.F. Tuan, Advances in Geophysics, Vol. 16, (1973)
- P.A. Sturrock, Astrophys. J. 182 569 (1973)













PENN STATE UNIVERSITY LIBRARIES



A000070945461